



# The SDSS-IV MaNGA Sample: Design, Optimization, and Usage Considerations

David A. Wake<sup>1,2,3</sup>, Kevin Bundy<sup>4,5</sup>, Aleksandar M. Diamond-Stanic<sup>2,6</sup>, Renbin Yan<sup>7</sup>, Michael R. Blanton<sup>8</sup>,  
Matthew A. Bershadsky<sup>2</sup>, José R. Sánchez-Gallego<sup>9</sup>, Niv Drory<sup>10</sup>, Amy Jones<sup>11</sup>, Guinevere Kauffmann<sup>11</sup>,  
David R. Law<sup>12</sup>, Cheng Li<sup>13,14</sup>, Nicholas MacDonald<sup>9</sup>, Karen Masters<sup>15,18</sup>, Daniel Thomas<sup>15,18</sup>,  
Jeremy Tinker<sup>8</sup>, Anne-Marie Weijmans<sup>16</sup>, and Joel R. Brownstein<sup>17</sup>

<sup>1</sup> School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA UK; [david.wake@open.ac.uk](mailto:david.wake@open.ac.uk)

<sup>2</sup> Astronomy Department, University of Wisconsin-Madison, Madison, WI 53706, USA

<sup>3</sup> Department of Physics, University of North Carolina Asheville, One University Heights, Asheville, NC 28804, USA

<sup>4</sup> Dept. of Astronomy and Astrophysics, UC Santa Cruz, MS: UCO/LICK, 1156 High St, Santa Cruz, CA 95064, USA

<sup>5</sup> Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

<sup>6</sup> Department of Physics and Astronomy, Bates College, 44 Campus Avenue, Carnegie Science Hall, Lewiston, Maine 04240, USA

<sup>7</sup> Department of Physics and Astronomy, University of Kentucky, 505 Rose Street, Lexington, KY 40506-0057, USA

<sup>8</sup> Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, NY 10003, New York, USA

<sup>9</sup> Department of Astronomy, Box 351580, University of Washington, Seattle, WA 98195, USA

<sup>10</sup> McDonald Observatory, University of Texas at Austin, 1 University Station, Austin, TX 78712-0259, USA

<sup>11</sup> Max-Planck Institut für Astrophysik, D-85741 Garching, Germany

<sup>12</sup> Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

<sup>13</sup> Department of Physics and Tsinghua Center for Astrophysics, Tsinghua University, Beijing 100084, China

<sup>14</sup> Shanghai Astronomical Observatory, Nandan Road 80, Shanghai 200030, China

<sup>15</sup> Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, UK

<sup>16</sup> School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK

<sup>17</sup> Department of Physics and Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112, USA

Received 2017 April 27; revised 2017 July 7; accepted 2017 July 7; published 2017 August 4

## Abstract

We describe the sample design for the SDSS-IV MaNGA survey and present the final properties of the main samples along with important considerations for using these samples for science. Our target selection criteria were developed while simultaneously optimizing the size distribution of the MaNGA integral field units (IFUs), the IFU allocation strategy, and the target density to produce a survey defined in terms of maximizing signal-to-noise ratio, spatial resolution, and sample size. Our selection strategy makes use of redshift limits that only depend on  $i$ -band absolute magnitude ( $M_i$ ), or, for a small subset of our sample,  $M_i$  and color ( $\text{NUV} - i$ ). Such a strategy ensures that all galaxies span the same range in angular size irrespective of luminosity and are therefore covered evenly by the adopted range of IFU sizes. We define three samples: the Primary and Secondary samples are selected to have a flat number density with respect to  $M_i$  and are targeted to have spectroscopic coverage to 1.5 and 2.5 effective radii ( $R_e$ ), respectively. The Color-Enhanced supplement increases the number of galaxies in the low-density regions of color-magnitude space by extending the redshift limits of the Primary sample in the appropriate color bins. The samples cover the stellar mass range  $5 \times 10^8 \leq M_* \leq 3 \times 10^{11} M_\odot h^{-2}$  and are sampled at median physical resolutions of 1.37 and 2.5 kpc for the Primary and Secondary samples, respectively. We provide weights that will statistically correct for our luminosity and color-dependent selection function and IFU allocation strategy, thus correcting the observed sample to a volume-limited sample.

**Key words:** galaxies: evolution – galaxies: general – galaxies: statistics – surveys

## 1. Introduction

The SDSS-IV MaNGA survey (Bundy et al. 2015; Blanton et al. 2017) is using the ARC 2.5 m telescope (Gunn et al. 2006) and the BOSS spectrographs (Smee et al. 2013) with its fibers bundled into multiple integral field units (IFUs; Drory et al. 2015) to measure spatially resolved spectroscopy of  $\sim 10,000$  nearby galaxies. We have chosen to target a well-defined sample that has uniform spatial coverage in units of  $r$ -band effective radius along the major axis ( $R_e$ ), and an approximately flat stellar mass distribution with  $10^9 \lesssim M_*/M_\odot h^{-2} \lesssim 10^{11}$ . In this paper, we discuss the motivation and methodology of the MaNGA sample selection, and we present the resulting sample in a way that allows for its use in statistical analysis of galaxy properties.

The challenge of designing a survey like MaNGA is to balance the need for sample size, spatial coverage, and spatial

resolution; these three parameters compete with each other for finite fiber resources. We have chosen a sweet spot in this multi-parameter space that best matches our science requirements (outlined in Bundy et al. 2015; Yan et al. 2016) in the context of a six-year survey duration, existing spectrographs, and telescope field of view. Since the sample design and the modifications to the BOSS spectrographs' fiber feeds (Drory et al. 2015) occurred concurrently, we were able to optimize both together to a considerable degree. Specifically, we determined the optimal IFU size complement within the confines of a total fiber budget and viable sample design. Fortunately, the redshift range  $0.02 \lesssim z \lesssim 0.1$  that balances angular size versus resolution also delivers a target surface density that is well matched to the telescope field of view (3 degrees in diameter) and the roughly 1500 fibers with 2" diameters of MaNGA's feed to the BOSS spectrographs. While we had not foreseen how well matched the telescope and instrument "grasp" were to our optimized target density, in

<sup>18</sup> SEPnet, South East Physics Network ([www.sepnet.ac.uk](http://www.sepnet.ac.uk)).

hindsight it is a lesson learned for planning future surveys. One of the aims of this paper is to demonstrate how, with adequate knowledge of target density, well-matched instrumentation can be optimally configured to achieve well-motivated survey science requirements.

A number of our design choices, such as an even sampling in stellar mass, roughly uniform radial coverage, and a sample size in the thousands, are similar in spirit to those of the SAMI survey (Croom et al. 2012; Bryant et al. 2015). Such choices result naturally from a desire to efficiently study the local galaxy population and produce several similar features in the sample selection approach, such as a stellar mass-dependent redshift range. However, our ability to simultaneously design the IFU size distribution and sample selection using a telescope with a larger field does offer further advantages for optimization.

### 1.1. Design Strategy

A number of strategic and tactical choices inform technical elements of the sample design. A starting point was to select from the well-understood SDSS Main Sample (Strauss et al. 2002) with enhanced redshift completeness and remeasured photometry, as described in Section 2. Because the redshifts and global properties of SDSS galaxies are well known, the distributions of these properties in the final MaNGA sample can be carefully constructed by effectively weighting the MaNGA selection in order to maximize its scientific utility.

1. Sample size: Paramount is the requirement for a large, statistically powerful sample size, a choice that comes at the expense of higher quality data for individual galaxies within the sample. As described in Bundy et al. (2015), the specific argument for sampling 10,000 galaxies arises from the desire to divide galaxies into  $6^3$  groups of  $\sim 50$  galaxies each. These groups, or bins (i) sample each of three “principal components” defining galaxy populations—stellar mass, SFR, and environment; (ii) divide each “dimension” into six bins, sufficient to distinguish the functional form of trends across each dimension; and finally (iii) contain adequate counting statistics (galaxies) such that differences in mean properties between bins can be detected at the five-sigma level even when the measurement precision for individual galaxies is comparable to this difference. This optimization dovetails MaNGA’s scientific goals for statistical analyses of resolved galaxy samples, and complements existing, smaller data sets such as ATLAS3D (Cappellari et al. 2011), DiskMass (Bershady et al. 2010), and CALIFA (Sánchez et al. 2012), as well as forthcoming data from instruments such as MUSE (Bacon et al. 2010) and KCWI (Martin et al. 2010) capable of producing even higher fidelity data for more modest samples.
2. Sampling in stellar mass: We desire the MaNGA sample to have a roughly flat distribution in  $\log M_*$  so that studies of mass-dependent trends could make use of adequate numbers of high-mass galaxies compared to more numerous low-mass systems. A flat stellar mass distribution requires an upper redshift limit that is stellar mass dependent, so a larger volume is sampled for rarer high-mass galaxies.
3. Radial coverage: We desire roughly uniform radial coverage as defined by some multiple of the effective

radius. This choice is motivated by the existence of well-known scaling relations that emphasize the importance of the relative length scale of galaxy stellar density profiles. Uniform spatial coverage in units of  $R_e$  requires a lower redshift limit that is stellar mass dependent, so larger more massive galaxies have the same angular size as smaller lower mass galaxies. MaNGA therefore samples the same relative extent of the declining surface brightness profile, but at the cost of *not* maintaining the same *physical* spatial resolution across the sample.

4. Maximize spatial resolution and signal-to-noise ratio (S/N): With current facilities, we wish to build a data set of IFU spectroscopy for 10,000 galaxies with the maximum possible per galaxy physical spatial resolution, spectral coverage and resolution, and S/N per spatial element. These requirements lead to several inevitable tactical features of the selection criteria:
  - (a) to maximize the spatial resolution and total S/N requires the selection of galaxies at as low redshift as possible.
  - (b) to reach our goal of 10,000 galaxies requires a sufficiently broad redshift distribution so as to have enough galaxies per plate to maximize efficiency in IFU allocation.

#### 1.1.1. Subsamples

The question of how to set the target radius motivated significant thought during the sample design. Smaller multiples of the effective radius would yield greater spatial resolution and more spatial samples with higher S/N. Larger radial coverage would contain more of the galaxy’s light, reach into the dark-matter-dominated regime, and probe uncharted territory in the outskirts of galaxies. After studying a number of options, a compromise was reached to cover out to  $1.5 R_e$  (the majority of the light distribution) for two-thirds of the sample and to cover out to  $2.5 R_e$  for one-third of the sample. Going to larger radii, while compelling, was deemed too costly in terms of the number of spatial samples per IFU with very low S/N. The sample split was motivated by basic binning arguments (see Bundy et al. 2015) and officially adopted by the science team after the first year of observations.

With the main sample roughly flat in  $\log M_*$ , it was possible to consider a further optimization, that is, balancing the rest-frame color distribution (a proxy for star formation rate) at fixed  $M_*$ . In this way, rare populations of star-forming massive galaxies and non-star-forming low-mass galaxies could be upweighted in the final sample. The primary objection was a concern that unexpected biases could be introduced into the sample and more generally that the selection would become unnecessarily complicated. As described below, a practical solution was discovered, however, that helps balance the color distribution through an additional and modest “Color-Enhanced supplement.” Should it prove biased or undesirable, the supplemental sample could be easily separated from the Primary sample, and in the worst-case scenario, even ignored. With the risk mitigated, the decision was made to include the Color-Enhanced supplement in the selection.

To summarize, the final full MaNGA sample with which we began the survey consists of three main subsamples. The Primary sample, which will initially make up 50% of the targets, is designed to be covered by our IFUs to  $1.5 R_e$  and has

a flat distribution in K-corrected  $i$ -band absolute magnitude ( $M_i$ ). The Secondary sample, making up 33% of the initial targets, is again designed to have a flat distribution in  $M_i$  but with coverage to  $2.5 R_e$ . Finally, the Color-Enhanced supplement is designed to add galaxies in regions of the  $\text{NUV} - i$  versus  $M_i$  color-magnitude plane that are under-represented in the Primary sample, such as high-mass blue galaxies and low-mass red galaxies, and will make up 17% of the initial targets. The combination of the Primary and Color-Enhanced samples is called the Primary+ sample.

This complexity leads to the final strategic choice in the survey design:

5. Selection simplicity: While we have described the basic strategic and tactical motivations behind various choices for the sample design, we were also driven to make the selection as simple and reproducible as possible. The implementation of the “weighting” described above to deliver a MaNGA sample with desired global distributions is carried out entirely through a set of selection criteria involving basic observables that are relatively model independent: redshift,  $i$ -band luminosity, and, for the Color-Enhanced supplement,  $(\text{NUV} - i)$  color. Note that the selection does not depend on effective radius explicitly (although a radius estimate is used when choosing what sized IFU to allocate to given galaxy target). We also emphasize that while much of the sample design studies made use of  $M_*$  estimates, the final selection employs  $i$ -band absolute magnitudes as a proxy for  $M_*$ .<sup>19</sup> We did not use  $M_*$  estimates specifically in order to avoid potential systematic biases and the use of a “black-box” estimator that may be difficult to reproduce.

### 1.2. Extant Instrumentation

Various aspects of the sample design are dependent on the nature of the MaNGA instrumentation. We highlight a few details here and refer to Drory et al. (2015) for more details.

The MaNGA instrumentation suite is composed of fiber-bundle IFUs dedicated to observing galaxy targets, with a number of additional IFUs and single fibers reserved for calibration. The total number of fibers, 1423, is limited by the size of the inherited BOSS spectrographs. The science IFUs contain circular, buffered optical fibers tightly arranged in a hexagonal format. This geometry enables IFUs of different sizes, with specific numbers of fibers for each IFU size. With a “live-core” fiber diameter of  $2''$  and full outer diameter of  $2''.5$ , the smallest of the science IFUs contains a central fiber and two outer, hexagonal rings for a total of 19 fibers and long-axis IFU diameter of  $12''.5$ . Other possible IFU sizes are 37 fibers ( $17''.5$ ), 61 fibers ( $22''.5$ ), 91 fibers ( $27''.5$ ), 127 fibers ( $32''.5$ ), 169 fibers ( $37''.5$ ), 217 fibers ( $42''.5$ ), and so on. Choosing the largest IFU size as well as the optimal distribution of IFU sizes is a major focus of this paper.

Identical sets of the MaNGA instrumentation suite are installed in six SDSS “cartridges.” These sturdy, cylindrical structures house the light-collecting IFUs and fibers, the field-specific plug plate, and the output pseudo-slit, which is directed

into the spectrographs when the cartridge is mounted on the telescope. The ferrules and jacketing of single fibers and MaNGA IFUs are similar, with dimensions that facilitate hand-plugging of these elements into pre-drilled plates. As a result there is a “collision radius” that defines the minimum distance between plugged elements. For the MaNGA IFUs this distance is  $120''$ . The mounting of the plate in the cartridge makes use of a post that attaches to the center of the plate helping to deform the plate to the shape of the focal plane. This center post introduces a second “collision radius” about the center of the plate of  $150''$ .

The balance of this paper is organized as follows: In Section 2 we describe the construction of the parent catalogs from the NASA-Sloan Atlas (NSA). In Section 3.1 we describe the process by which we select the upper and lower redshift cuts for our Primary and Secondary samples. In Section 3.2 and Section 3.3 we describe the methodology that we use to optimize the IFU size distribution. In Section 3.4 we describe how we select the sample space density. In Section 4 we describe the results of applying these processes, the selection of the Color-Enhanced Supplement, and the properties of the final samples. In Section 5 we describe how we tile the survey area and allocate IFUs to the targets. In Section 6 we discuss how to use the sample for statistical analyses of MaNGA data.

Where applicable we use a flat  $\Lambda$ -CDM cosmology with  $\Omega_M = 0.3$  and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  except for absolute magnitudes and stellar mass, which are calculated assuming  $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$  with  $h = 1$ , following previous versions of the NSA.

## 2. Parent Catalogs

The primary input for the selection of all MaNGA galaxies is an enhanced version of the NSA (Blanton M. <http://www.nsatlas.org>). The NSA is a catalog of nearby galaxies within 200 Mpc ( $z \simeq 0.055$ ), primarily based on the SDSS DR7 MAIN galaxy sample (Abazajian et al. 2009), but incorporating data from additional sources. The SDSS imaging has been reprocessed to be better suited to the analysis of these large nearby galaxies (Blanton et al. 2011). In particular it has improved background subtraction and deblending more suited to nearby large galaxies resulting in more accurate size and luminosity measurements for such galaxies. In addition to a reanalysis of the SDSS imaging, a similar analysis is applied to the *GALEX* near- and far-UV images, and several derived parameters, such as K-corrections and absolute magnitudes<sup>20</sup> (using *kcorrect* v4\_3), Sérsic profile fits, and stellar masses are determined.

The NSA also provides a  $\sim 30\%$  improvement in spectroscopic completeness over the standard SDSS spectroscopic catalog for the very brightest sources by adding redshifts from the NASA Extragalactic Database (NED<sup>21</sup>), the CfA Redshift Survey (ZCAT;<sup>22</sup> Huchra & Geller 1991), the Arecibo Legacy Fast ALFA Survey (ALFALFA; Giovanelli et al. 2005), the 2dF Galaxy Redshift Survey (2dF; Colless et al. 2001), and the 6dF Galaxy Redshift Survey (6dF; Jones et al. 2009). The SDSS is 70% complete at  $r_{AB} \sim 14$  and 95% complete at  $r_{AB} \sim 16$ , emphasizing the importance of these other redshift

<sup>19</sup> For the initial IFU size distribution optimization process we used the stellar mass as estimated by the *kcorrect* code (Blanton & Roweis 2007) applied to the five-band SDSS photometry. For the final samples we have switched to using just  $i$ -band absolute magnitude in order to simplify the selection function (see Section 4.1).

<sup>20</sup> K-corrections in the NSA catalog do not account for extinction explicitly, and make no attempt to apply an inclination-dependent extinction correction.

<sup>21</sup> <https://ned.ipac.caltech.edu/>

<sup>22</sup> <https://www.cfa.harvard.edu/~dfabricant/huchra/zcat/>



sources. Assuming that the incompleteness of SDSS is orthogonal to the incompleteness of the other sources, we estimate that the completeness of the combined sample between  $13 < r_{AB} < 17$  is 98.6%.

In order to achieve our primary sample design goals (radial coverage and stellar mass range), we need to target massive galaxies at  $z > 0.055$ . We have thus extended the NSA analysis to include galaxies with  $z < 0.15$ .

We have made one further addition to the standard NSA analysis, the calculation of elliptical Petrosian magnitudes and profiles for all seven bands. The elliptical Petrosian method uses a set of elliptical annuli defined using an estimate of the axis ratio  $b/a$  (minor over major) and the position angle  $\phi$ . Otherwise it uses the standard algorithm for Petrosian magnitudes, with the Petrosian radius  $r_p$  defined as the major axis of the ellipse where the Petrosian ratio  $\eta = 0.2$ , and with the aperture for the flux defined with a major-axis radius  $2r_p$ .  $R_e$  is defined as the major axis radius of the ellipse that contains 50% of the flux within  $2r_p$ . The NSA pipeline produces several estimates of  $b/a$  and  $\phi$  but for the elliptical Petrosian method we use those determined using the second moments within the circularized Petrosian  $r_{90}$ , the radius containing 90% of the flux within  $2r_p$ .

We have also applied aperture corrections to the photometry, to account for the variation in point-spread function between the bandpasses, particularly for *GALEX*. We do so by using the measured curve-of-growth to predict the aperture correction for an ideal elliptical galaxy, and applying this correction to the real data. For *GALEX*, these corrections can be of the order of 30% to 50% for galaxies with half-light radii around an arcsecond; for SDSS they are always negligible.

Visual inspection of our targets during the target selection process revealed that the Sérsic profile fitting suffered more catastrophic failures than the circular Petrosian profile calculation. A detailed comparison with the Simard et al. (2011) two-component GIM2D Sérsic fits further showed that the NSA's single Sérsic  $R_e$  estimates are systematically overestimated for early-type galaxies (or galaxies with high concentrations) by up to 50% at high Sérsic- $n$ . Adding the elliptical Petrosian fitting maintained the stability of the circularized fits while also measuring axis ratios and position angles, and they do not show systematic differences with the two-component Sérsic fits. We therefore choose to use elliptical Petrosian  $R_e$  and flux measurements throughout. Absolute magnitudes and stellar mass in the NSA, and hence used here, are calculated assuming  $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$  with  $h = 1$ .

This extended NSA is designated v1\_0\_1 and is publicly available as part of the SDSS data releases from DR13 onward.

For the MaNGA selection, we limit the extended NSA catalog to those galaxies that lie within the Large Scale Structure mask produced as part of the Data Release Seven NYU Value Added Galaxy Catalog (Blanton et al. 2005). This ensures that all targets fall in regions with good SDSS photometry and spectroscopic coverage, and are not close to very bright stars.

### 3. Constructing the Targeting Catalog

Given the parent catalogs defined above, we now discuss the construction of the “targeting catalog” that will define the final selection from which the MaNGA targets will be allocated. We are guided by three requirements.

1. More than 80% of the sample should have a specified radius (e.g., 1.5 or 2.5  $R_e$ ) smaller than our largest IFU bundle.<sup>23</sup>
2. A flat distribution in the stellar mass proxy with a low-mass limit of  $\sim 10^9 M_\odot$ .
3. The selection will only use cuts in redshift that depend on the stellar mass proxy (and one color in the case of the Color-Enhanced supplement).

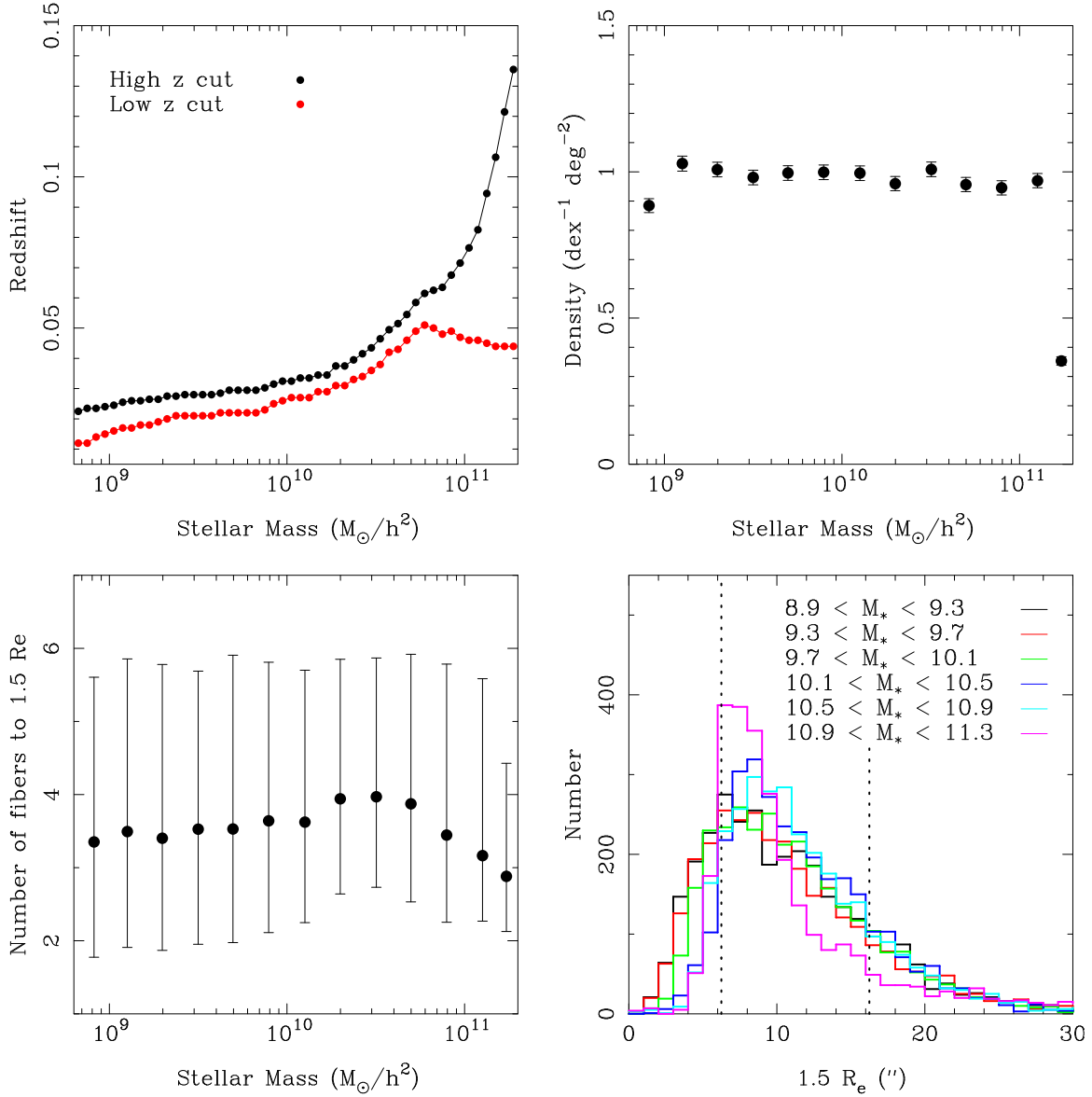
A summary of the targeting catalog construction is as follows. We consider three targeting samples. The goal of the Primary sample is to provide coverage to a radius corresponding to 1.5  $R_e$ . The Color-Enhanced supplement (roughly 17% of the final sample) produces a more uniform coverage in  $\text{NUV} - i$  color as a function of mass when combined with the Primary sample to form the Primary+ sample. The Secondary sample, designed to yield a sample size that is half of the Primary+ sample, covers larger radii, up to 2.5  $R_e$ . For the optimization process that we describe below we only consider the Primary and Secondary samples. The Color-Enhanced supplement is produced by only slightly widening the Primary sample selection criteria in a color-dependent way. That combined with its small size means that it has a negligible effect on the final sample size and S/N distributions and so the optimization based on the Primary and Secondary samples remains valid (see Section 4.5 for a demonstration of this and a detailed description of the Color-Enhanced selection methodology).

After choosing the relative proportions of the subsamples, we first adopt a desired total sky density of potential targets. This in itself requires an optimization process that balances the efficiency of allocating IFUs, the field of view, survey area, and the number and size of IFUs that can be constructed, and trade-offs in S/N, exposure time, spatial resolution, and radial coverage. These are discussed in Section 3.4. Once the desired sky density is defined, we derive stellar mass proxy dependent low- and high-redshift cuts that yield samples that meet the coverage criteria. These cuts are then optimized to deliver the highest S/N and spatial resolution across the targeting samples (in effect, this means that the lowest redshifts are preferred). We then “tile” the survey—a term that refers to the selection of MaNGA pointings across the sky and the allocation of IFUs to targets. This allows us to evaluate the final “observed” sample that is obtained as well as the frequency of unused or improperly allocated IFU bundles. We repeat the process multiple times under different assumptions for the target density, the minimum and maximum IFU sizes, and the distribution of fabricated IFU sizes to determine the optimal configuration. Further details are given below.

#### 3.1. Selecting Upper and Lower Redshift Cuts

Once the desired sky density has been set (see Section 3.4), we identify redshift intervals at every stellar mass where  $>80\%$  of galaxies with that mass have a physical scale (either 1.5 or 2.5  $R_e$ ) that subtends an angular size that fits within the largest available IFU (for discussion on the maximum IFU size, see Section 3.3). There are many such redshift intervals, of course. By choosing the interval with the lowest redshift we maximize both the spatial resolution (in kiloparsecs) and the S/N of the resulting sample, while maintaining both the radial coverage

<sup>23</sup> We define the radius of our hexagonal IFUs to be the radius of the circle that has the same area as the IFU.



**Figure 1.** Demonstration of the sample selection methodology. The top left panel shows the lowest redshift interval at each stellar mass that will produce a sample of galaxies where 80% can be covered to  $1.5 R_e$  by the 127 fiber IFU (the largest available in this simulation) with a flat number density distribution as a function of stellar mass with a density of  $1 \text{ deg}^{-2} \log(M_*)^{-1}$ . The top right panel shows the resulting stellar mass distribution when these cuts are applied to the parent catalog. The bottom left panel shows the distribution of the number of  $2''$  5-spaced fibers that are required to cover  $1.5 R_e$  of each galaxy. The points show the median and the error bars show the 20th and 80th percentiles. The central fiber is not counted, i.e., the 127 fiber IFU has 6 radial fibers. The bottom right panel shows the resulting angular size distribution of  $1.5 R_e$  in bins of stellar mass. The distributions are very similar regardless of stellar mass, with the exception of the most massive galaxies. The vertical dotted lines show the radii of the 19 and 127 fiber IFUs.

and density criteria. We impose a hard lower redshift limit of  $z = 0.01$ , designed to minimize the distance errors introduced by the local velocity field. This lower redshift limit also has the effect of limiting the main samples to stellar masses larger than  $\sim 4 \times 10^8 M_\odot h^{-2}$ .

In practice, we bin the parent catalog into a fine grid in log stellar mass (or absolute magnitude) and redshift. For each stellar mass bin we find all the redshift ranges that produce the target density. We then find the lowest redshift range that yields a sample in which 80% of galaxies can be covered to  $1.5$  or  $2.5 R_e$  (for the Primary and Secondary samples, respectively). This produces an upper and lower redshift limit for each stellar mass bin. We then interpolate to find the appropriate redshift thresholds for every stellar mass in the parent catalog.

Figure 1 shows the results of applying this method under the assumption of a sky density of  $1 \text{ deg}^{-2} \log(M_*)^{-1}$  and a maximum IFU size of 127 fibers. While the upper and lower redshift cuts (top left panel) may look somewhat convoluted, they produce a sample that has a very similar angular size distribution across all stellar masses. This means that we probe the same spatial resolution in units of  $R_e$  at all masses, although the physical resolution in kiloparsecs is mass dependent. We will return to this point later. We note that the change in the distributions for the highest mass galaxies is unavoidable. Due to the steepness of the mass function there is a shortage of these galaxies at very low redshifts.

The bottom panels of Figure 1 also show that even in a narrow mass and redshift range the galaxies show a wide variation in size. In fact even if we were to look at a single

stellar mass (or magnitude) and redshift, there is still a significant range in galaxy  $R_e$  (rms  $\sim 50\%$ ). Therefore, a range in IFU sizes is required to most efficiently observe the sample.

### 3.2. Optimizing the IFU Size Distribution

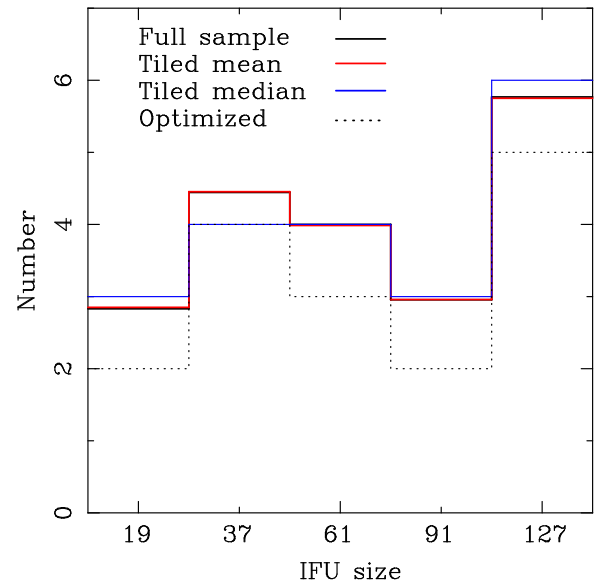
Physical size and mass are correlated 1:1 (to first order), and we wish to sample a factor of 100 in mass. To match this dynamic range requires either a comparable range in IFU size (for a pure, volume-limited sample), or selecting more massive galaxies at preferentially higher redshift, thus lowering physical resolution in a mass-dependent way. As the dynamic range in IFU size increases, the total number of IFUs decreases (for fixed total fiber number). This decrease in the number of IFUs then requires a lower target density, a larger survey footprint, and, for a fixed number of targets, shorter exposures.

To properly optimize these trade-offs we need to investigate the final sample properties after a full realization of our targeting and selection algorithms, rather than just analyzing the targeting catalog itself. In particular we must include the effects of the field size (i.e., we must tile the survey) and available IFUs. We first tackle the question of how to optimize the IFU size distribution given a maximum and minimum IFU size. In the next section we will discuss how the maximum and minimum size is chosen.

Since our final sample of galaxies will have a range of apparent angular sizes (see Figure 1) we would like to have a range of IFU sizes that is able to closely match this distribution. We always wish to observe a galaxy with an IFU that is large enough to reach the desired radius, but, to maximize survey efficiency, we do not wish to use an IFU any larger. If in a given tile<sup>24</sup> we have many more targets than available IFUs, we can select the galaxies that match our IFU size distribution and thus maximize our survey efficiency. However, if the IFU distribution does not match the underlying galaxy size distribution, we will produce a final sample that is biased compared to our input sample. For these reasons we want to carefully select the IFU size distribution that most closely matches the per tile size distribution derived from the targeting catalog.

Figure 2 shows the distribution of required IFU sizes for a potential targeting catalog. The construction of this catalog assumed a Primary-to-Secondary ratio of 2–1, a maximum IFU size of 127 fibers, and a Primary sample density of  $1 \text{ deg}^{-2} \log(M_*)^{-1}$ . Galaxies that require an IFU size smaller than 19 fibers are assigned to a 19 fiber IFU, and likewise, galaxies that require an IFU size larger than 127 fibers are assigned a 127 fiber IFU. The black histogram in Figure 2 shows the size distribution derived from the full targeting catalog, assuming that *all* galaxies can be equally well observed. In fact, galaxies can “collide” (if a pair is closely separated only one may be allocated an IFU) and are highly clustered, resulting in some regions with more or fewer available targets than MaNGA has IFUs.

After running a non-overlapping tiling of the targeting catalog (see Section 5), the importance of these effects can be judged by the resulting mean (red histogram) and median (blue histogram) distributions, defined *per tile*. In all cases, the size



**Figure 2.** Required IFU size distribution to cover the primary sample to  $1.5 R_e$  and the Secondary sample to  $2.5 R_e$ . Galaxies smaller than 19 fibers are assigned to 19 fiber IFUs; galaxies larger than 127 fibers are assigned to 127 fiber IFUs. The solid black histogram indicates the mean size distribution of the whole sample. The red histogram shows the mean size distribution per tile after allocating IFUs using a non-overlapping tiling of the sample. The blue histogram shows the median distribution per tile. All three solid histograms have been normalized to a total of 20 galaxies. The dotted histogram is the optimal IFU size distribution that can fit on the slit. See the text for the exact optimization procedure.

distributions are scaled to a total number of 20 IFUs, which is the median number of target galaxies per tile in the adopted tiling scheme. The red or blue histogram in comparison to the black histogram represents two extreme tiling strategies. The non-overlapping tiling is the most efficient in terms of maximizing the number of galaxies observed with a given number of plates while still probing all environments, but makes no attempt at completeness. The distribution in the full targeting catalog represents 100% completeness while paying no attention to efficiency. Our eventual strategy will be somewhere between the two, but we can see that there is practically no difference between the two means (black versus red).

Since we are limited to selecting integer numbers of IFUs (the blue histogram in Figure 2) the median distribution of the non-overlapping tiling, looks to be a good solution. However, these IFUs would require more fibers than can fit on the slit of the BOSS spectrograph even with our minimum acceptable slit spacing (see Drory et al. 2015). Therefore, we must find the IFU size distribution that most closely matches these required distributions but requires fewer fibers than can fit on the BOSS spectrograph slit.

To achieve this optimization we perform an exhaustive search over a large number of IFU size distributions where the number of IFUs of a given size varies from 0 to 8 and where the total slit space consumed is always less than the maximum available. We then calculate the mean square difference between each test IFU size distribution and the actual required IFU size distribution, where both are normalized to a total number of 20 IFUs. We do this both for the full targeting catalog size distribution and for each of the non-overlapping tiles. In the case of the non-overlapping tiles, the mean square difference is then summed over all tiles. The *best* IFU size

<sup>24</sup> We adopt the SDSS terminology that defines a pointing of the Sloan 2.5 m field of view on the sky as a “tile.” A given plate is associated with a set of drill holes that locate fibers on specific targets. Thus, more than one plate can be observed over a given tile. A full set of tiles, which may also overlap, describes the footprint of the survey.

distribution is then that which minimizes the mean square difference.

For the sample described in this section the optimal IFU size distribution is 2, 4, 3, 2, 5 for both the tiled and untiled samples. It is shown as the dotted line in Figure 2. We note that a distribution of 2, 4, 4, 2, 5 is almost as good a fit and can also be accommodated on the slit. Since it wholly contains the optimal distribution but includes an extra IFU it would be logical to choose this distribution as it will yield a larger final sample that can be reduced to the optimal distribution (2, 4, 3, 2, 5) after the observations are completed, if so desired.

### 3.3. Selecting the Maximum and Minimum IFU Size

Early work defining the survey and instrument strategy resulted in the definition of a sample that required an IFU size range from 19 to 127 fibers. This size range was determined by a combination of the requirements of the initial sample selection and the properties of the instrument and has been used for the IFU development work. In this section we revisit this choice of IFU size range and investigate if it is optimal.

#### 3.3.1. Minimum IFU Size

The choice of the the 19 fiber IFU as the minimum possible size is a simple one. We require at least three radial bins for all of our science cases and so smaller IFUs are not worthwhile. Our sample selection methodology (Section 3.1) is not directly constrained by the minimum IFU size, but instead maximizes the angular size of the sample. As such, we can make the 19 fiber IFU available to our IFU size distribution optimization procedure (Section 3.2) and see if it is required for a given sample.

#### 3.3.2. Maximum IFU Size

The choice of a maximum IFU size is somewhat more complex as it enters directly into the determination of the stellar mass-dependent redshift cuts that define our samples (Section 3.1). A larger maximum IFU size will allow a sample to have a lower average redshift and still be covered to the same physical scale (e.g., 1.5 or 2.5  $R_e$ ). Clearly selecting a sample at lower redshift will improve the resolution and could potentially increase the S/N. The downside is that an increase in the IFU size reduces the total number of IFUs that will be available, since the slit length and hence number of fibers is fixed. Thus, to achieve the same sample size with fewer IFUs the number of plates observed must increase, and thus with a fixed amount of survey time available the exposure time per plate must be reduced.

We investigate these trade-offs by designing a series of samples with different maximum IFU sizes of 91, 127, 169, 217. In each case we design Primary and Secondary samples where 80% of the galaxies are covered to 1.5 and 2.5  $R_e$ , respectively, by the maximum IFU size. We choose the target densities of each sample so that when they are tiled each sample has the same fraction of IFUs that are unused due to tiles with too few targets on them. For each sample, we optimize the IFU size distribution using the procedure described in Section 3.2 utilizing a full non-overlapping tiling, with the results shown in Table 1.

We then tile each of these samples using the optimized IFU distribution and a non-overlapping tiling, selecting the number of tiles required to produce a sample of 9000 galaxies. For each

tile, galaxies are selected to match the available IFU sizes. If there are more IFUs of a given size than galaxies of that size, they are allocated to galaxies that cannot be allocated an IFU of the correct size (see Section 5.3 for more details of the allocation process).

Table 2 gives some of the properties of these samples. Details of the performance of the Primary sample are shown in Figure 3. The S/N values in the table and plots are calculated using an exposure time inversely proportional to the required number of plates. We assume an exposure time of 3 hr for the sample designed for a maximum IFU size of 127 fibers and scale the exposure time for the other samples accordingly to ensure the same total survey time.

The top row of panels in Figure 3 simply shows that the Primary samples are performing as designed. The center row compares the S/N properties and the bottom row the resolution and coverage. We assume an effective angular resolution of 2''5 based on expectations for the reconstructed data cubes. The center left panel shows that the median S/N per fiber at 1.5  $R_e$  decreases as the maximum IFU size increases. This S/N decrease is simply due to the decrease in exposure time per plate, required since we must observe more plates to achieve the same final sample size with fewer IFUs per plate. A better representation of the S/N is given by the two other panels in the center row, which show the S/N at 1.5  $R_e$  per kpc<sup>2</sup> and per  $R_e^2$ , respectively.<sup>25</sup> The opposite trend is now apparent, with the S/N increasing as the maximum IFU size increases, since each kiloparsec or unit of  $R_e$  covers more fibers at lower redshift. This trend is confirmed by the median S/N values for the whole of each sample given in Table 2. The largest fractional S/N increase occurs when increasing the maximum IFU size from 91 to 127 fibers, with the relative size of the increase diminishing as the maximum IFU size increases beyond 127.

As we allow larger IFUs to be considered the resolution increases as expected. Once again the biggest improvement is seen when going from a maximum IFU size of 91 to 127 fibers. This trend is not surprising since we increase the radius by one fiber each time and so there is a larger fractional increase at smaller IFU sizes. What is less obvious is the strong stellar mass dependence to the gain in resolution, with the largest effect occurring at stellar masses of a few  $10^{10} M_\odot$ . This reflects the fact that galaxies in this stellar mass range have the largest mean angular sizes because they are intrinsically larger than low-mass galaxies, but the turnover of the mass function means higher mass galaxies must be targeted at increasingly more distant redshifts.

The final panel (bottom right) shows the fraction of galaxies that are covered to at least 1.5  $R_e$  after tiling. There is a general slow reduction in this fraction as the maximum IFU size increases, but a sudden fall from 169 to 217. The fraction of the IFU complement made up by the largest IFU bundle size does decrease as the maximum allowed size increases, leading to a reduction in the fraction covered to the target radius. There are also fewer galaxies per plate for a given IFU size, making it harder to match the galaxies to the IFUs. This could be

<sup>25</sup> The S/N per  $R_e^2$  at 1.5  $R_e$  is a useful if perhaps unusual metric. To compute it, we consider the integrated S/N in a small annulus (or fiber) positioned at 1.5  $R_e$ . We then divide by the area of that annulus in units of  $R_e^2$ . This naturally accounts for the fact that the angle subtended by  $R_e$  on the sky (e.g., in arcseconds) depends on the galaxy's intrinsic size and its redshift. Furthermore, this S/N metric is appropriate for addressing the fidelity of measurements of both kinematics and compositional properties that scale with  $R_e$ .



**Table 1**  
Optimal IFU Size Distributions

Max IFU Size	N19	N37	N61	N91	N127	N169	N217	Mean Square Difference/dof
91	4	6	5	7	0	0	0	5584
127	2	4	4	2	5	0	0	5794
169	1	2	3	2	1	4	0	7072
217	1	2	2	1	1	1	3	6182

mitigated with a higher density sample but there would be a subsequent loss of resolution (see Section 3.4 below).

While increasing the maximum IFU size from 127 to 169 will lead to some gain in S/N and resolution (it should be noted that the same effect causes both to increase) with a moderate loss of coverage fraction, there are some more practical disadvantages to larger IFU sizes. Since larger IFUs require more slit space, the total number per plate decreases. This results in a larger number of plates required for the same sample size and hence a higher plate production cost (30% more plates  $\sim$  \\$100 k). Furthermore, building and testing an IFU larger than the 127 fiber IFU would have required further development, again increasing costs and placing the schedule at risk. We therefore settled on a final IFU size range of 19–127 fibers.

### 3.4. Choosing the Sample Density

It is possible to construct targeting catalogs meeting our science specifications with different number densities on the sky. Given that the spatial density of galaxies varies on the sky, a higher density of potential targets allows more efficient tiling and more efficient use of IFUs. A higher density can be achieved by widening the redshift intervals at a given stellar mass. While the average redshift would remain roughly constant, as required by the desire for a constant angular coverage, a wider redshift interval would result in a wider spread in angular sizes, increasing the tension between the dynamic range of galaxy sizes and the dynamic range of IFU sizes. In addition, as the desired sky density is increased, if a hard lower redshift limit (e.g.,  $z = 0.01$ ) is reached or there are too few massive galaxies at low- $z$ , the average redshift must be increased, resulting in poorer spatial resolution and total S/N.

Conversely, higher density samples have the advantage of requiring fewer plates with unused IFUs allowing us to reach the desired sample size of the main samples with fewer plates. Since our total time is fixed we may increase the exposure time and thus potentially the S/N.

Overall, input samples with higher density may have a slightly higher S/N (or more galaxies) but at a potential cost of lower spatial resolution and greater sample variance in both spatial resolution and S/N.

To investigate this trade-off, we generate several samples using the same procedure as described in Section 3.1 with a large range in sky density and  $M_i$  as our stellar mass proxy. These samples are then tiled using a 2, 4, 4, 2, 5 IFU size distribution and a non-overlapping tiling, adding tiles until a sample of 10,000 galaxies is reached. A Secondary sample is also included that has 50% of the density of the Primary sample.

Figure 4 shows the properties of these Primary input samples constructed from targeting catalogs with varying densities. The top left panel simply shows how the fraction of unused IFUs

depends on density, showing a rapid decline as density increases, which asymptotes to zero at high densities as expected. Reading from left to right and top to bottom, the next three panels show the density dependence of minimum, maximum and median redshift for all galaxies (black) and split by stellar mass (color). One can see that as the density increases the minimum and maximum redshifts diverge as expected. It is also evident that the median redshift increases little, except where  $z_{\min}(M_i)$  hits a limit, which is most evident for the highest and lowest stellar mass bins.

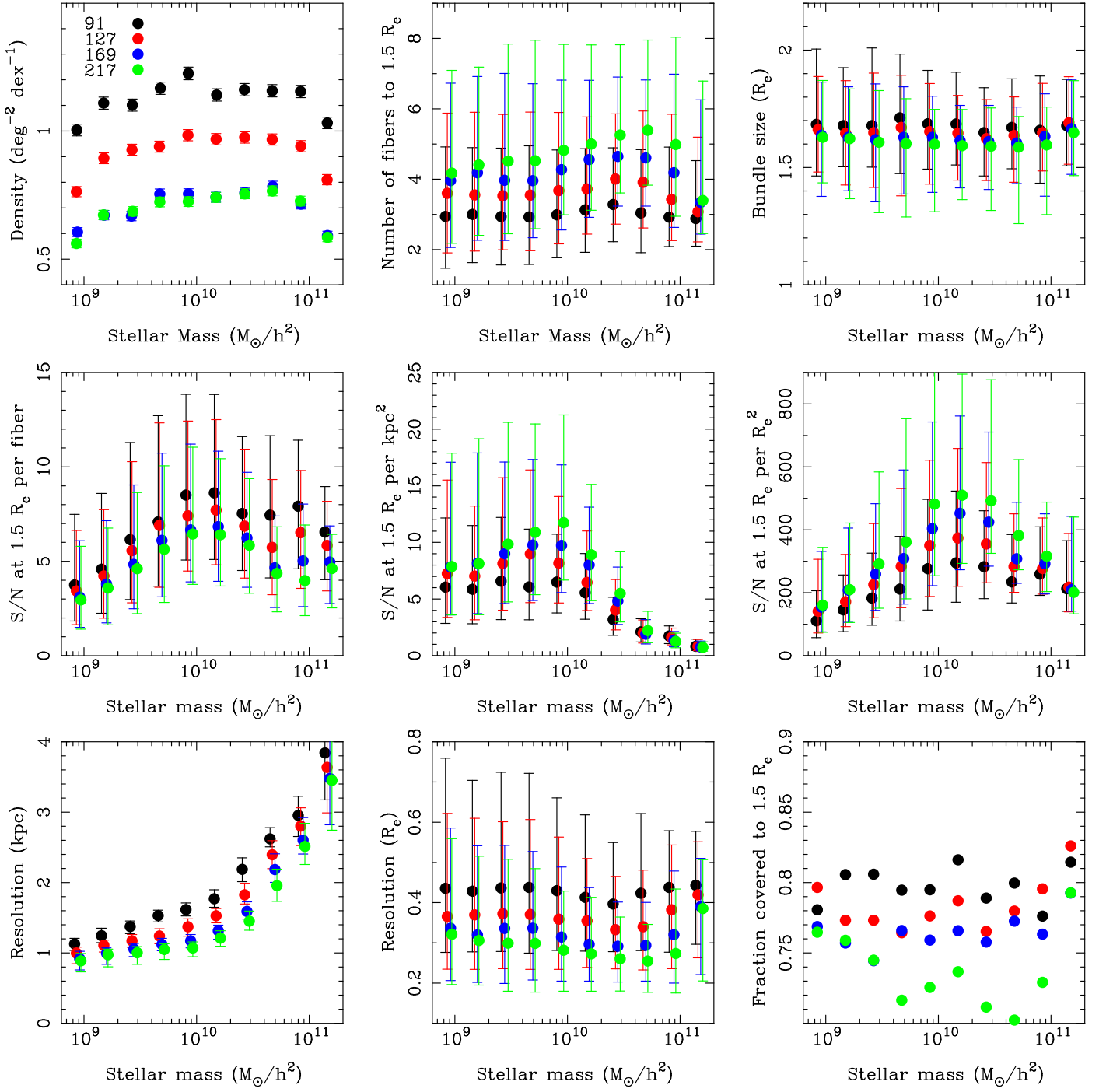
The final four panels show the median and rms of the S/N and resolution, respectively. The S/N is determined in a fiber at  $1.5 R_e$  and is divided by the area of the fiber in units of  $R_e^2$ . Likewise the resolution is given in units of  $R_e$ . For each density the S/N is scaled by the square root of the relative number of plates required to reach 10,000 galaxies, representing the change in plate exposure time available in a fixed-duration survey. One can see that as the density is increased the median S/N begins to increase before it reaches a plateau or turns over and begins to decrease. This turnover happens most rapidly for the lowest and highest mass samples reflecting their larger changes in median redshift, which counteracts the increased exposure time and decreases the S/N. The median resolution again shows the largest trend for the highest and lowest mass samples as it simply tracks the median redshift and thus increases (degrades) with density. In both cases the rms increases with density, reflecting the widening high- and low-redshift limits.

Figure 4 makes it clear that we do not wish to target samples with densities  $> 0.6 \text{ deg}^{-2} \text{ mag}^{-1}$  since the S/N has either flattened off or is declining at this point while the resolution gets poorer and the scatter in both quantities increases. However, at densities below this there is a trade-off between S/N and resolution. A density of  $0.53 \text{ deg}^{-2} \text{ mag}^{-1}$  maximizes the overall median S/N while only reducing the median resolution by 1% over the whole sample and produces similar results for the individual stellar mass bins with the exception of the highest stellar masses. We therefore select this density for the Primary sample.

## 4. Final Targeting Catalogs

We have described above our procedure and optimization strategy for constructing targeting catalogs for our Primary and Secondary samples. We have decided to allow IFU sizes of 19, 37, 61, 91, and 127 fibers and a density for the Primary+ sample of  $0.53 \text{ deg}^{-2} \text{ mag}^{-1}$ . If we wish to have a Secondary sample of 50% the size of the Primary+ sample, we would require a density of  $0.37 \text{ deg}^{-2} \text{ mag}^{-1}$ . This is not simply a factor of two lower than the Primary+ density since the  $M_i$  completeness limit of the extended NSA (Equation (4) below) means that we cover a narrower  $M_i$  range in the Secondary sample ( $M_i - 5 \log h \lesssim -18$ ) than in the Primary sample





**Figure 3.** Performance comparison of different samples designed with different maximum IFU size. Black, red, blue, and green symbols represent samples with a maximum IFU size of 91, 127, 169, and 217 fibers, respectively. See the text for detail.

( $M_i - 5 \log h \lesssim -19$ ). However, we have chosen to design the Secondary sample with a somewhat higher density of  $0.5 \text{ deg}^{-2} \text{ mag}^{-1}$ . The higher density increases the redshift range somewhat making the sample less sensitive to cosmic variance and thus reduces the plate-to-plate variation in the number of targets, making it easier to allocate IFUs to galaxies with the correct size. To maintain the desired 2:1 ratio of Primary+ to Secondary galaxies we down-sample the Secondary sample to a density of  $0.37 \text{ deg}^{-2} \text{ mag}^{-1}$  during the IFU allocation process.

#### 4.1. Simplifying the Selection Function

Since we have elected to design samples that have flat number densities as a function of a stellar mass proxy (and are flattened in color as well in the case of the Color-Enhanced supplement) we will need to correct for this imposed selection function for any statistical analysis of the MaNGA sample. Since the only selection we impose is an upper and lower redshift limit as a function of our stellar mass proxy (or color and mass for the Color-Enhanced supplement) we can exactly define the volume over which any galaxy in our samples could

**Table 2**  
The Properties of Primary Samples Designed for Differing Max IFU Sizes

Max IFU Size	Density (deg <sup>-2</sup> log( $M_*$ ) <sup>-1</sup> )	$N_{\text{gals}}$	$N_{\text{IFU}}$	$N_{\text{plates}}$	Fraction of Unused IFUs	Median S/N per kpc <sup>2</sup> at 1.5 $R_e$	Median S/N per $R_e^2$ at 1.5 $R_e$
91	1.2	9008	22	474	0.14	3.8	220
127	1	9006	17	596	0.11	4.6	268
169	0.8	9005	13	802	0.14	5.2	304
217	0.8	9003	11	915	0.11	5.7	344

have been selected. This allows the easy calculation of a  $V_{\text{max}}$  weight for every galaxy in the sample enabling the sample to be corrected back to the volume-limited case (see Section 6.1 for details).

However, this  $V_{\text{max}}$  is only perfectly defined in the case where there are no errors on the selection parameter (e.g., mass, magnitude, color etc.). Larger errors in the selection parameter, or the combination of errors on multiple selection parameters, translates to a larger error in the weight calculation, which would then translate into errors in any derived relations defined from the MaNGA sample. This could be particularly troubling if the error in the weight ended up correlating with an interesting derived parameter from the MaNGA IFU data.

For this reason we have elected to define the Primary and Secondary samples using just  $M_i$  rather than a full estimation of the stellar mass, and the Color-Enhanced supplement using  $\text{NUV} - i$  color rather than an estimate of the star formation rate. This is also the reason why we have separated the Primary and Color-Enhanced samples as we have, rather than making the Primary sample fully flattened in both a mass and an SFR proxy. Such a split allows the use of just the Primary sample for analyses that may be particularly sensitive to errors in the weights, but still enables analyses that will need better statistics in the lower density regions of the SFR–mass plane.

Flattening the density in  $M_i$  rather than mass has a disadvantage in which we end up with a significant number of low-mass ( $10^8$ – $10^9 M_\odot h^{-2}$ ) blue galaxies in the sample. We therefore choose to remove these by including a color-dependent absolute magnitude limit at the faint end. While this does not produce a hard cut in stellar mass, it does remove most of the very low-mass blue galaxies. These cuts are defined as

$$g - r > 0.4(M_r - 5 \log h) + 7.4 \quad (1)$$

and

$$g - r > 0.28(M_r - 5 \log h) + 5.6 \quad (2)$$

for the Primary and Secondary samples, respectively, where  $g - r$  is the k-corrected color at redshift zero.

The final simplification we make is to define the upper and lower redshift limits for the Primary and Secondary samples using a functional form rather than an interpolation between narrow redshift bins. This is largely done for ease of communication and reproduction and only results in a minor change in the sample properties and minimal impact on sample performance. These limits are defined by the following functional form:

$$z_{\text{lim}} = (A + B(M_i - 5 \log h))(1 + \exp[C(M_i - 5 \log h - D)]) \quad (3)$$

with parameters  $A$ ,  $B$ ,  $C$ , and  $D$  for the lower and upper redshift limits of each sample given in Table 3.

We also include a completeness cut required as a result of the magnitude limit of the input catalog where

$$z < -0.9335 - 0.1839(M_i - 5 \log h) - 0.01222(M_i - 5 \log h)^2 - 2.7668 \times 10^{-4}(M_i - 5 \log h)^3. \quad (4)$$

#### 4.2. Final Sample Properties

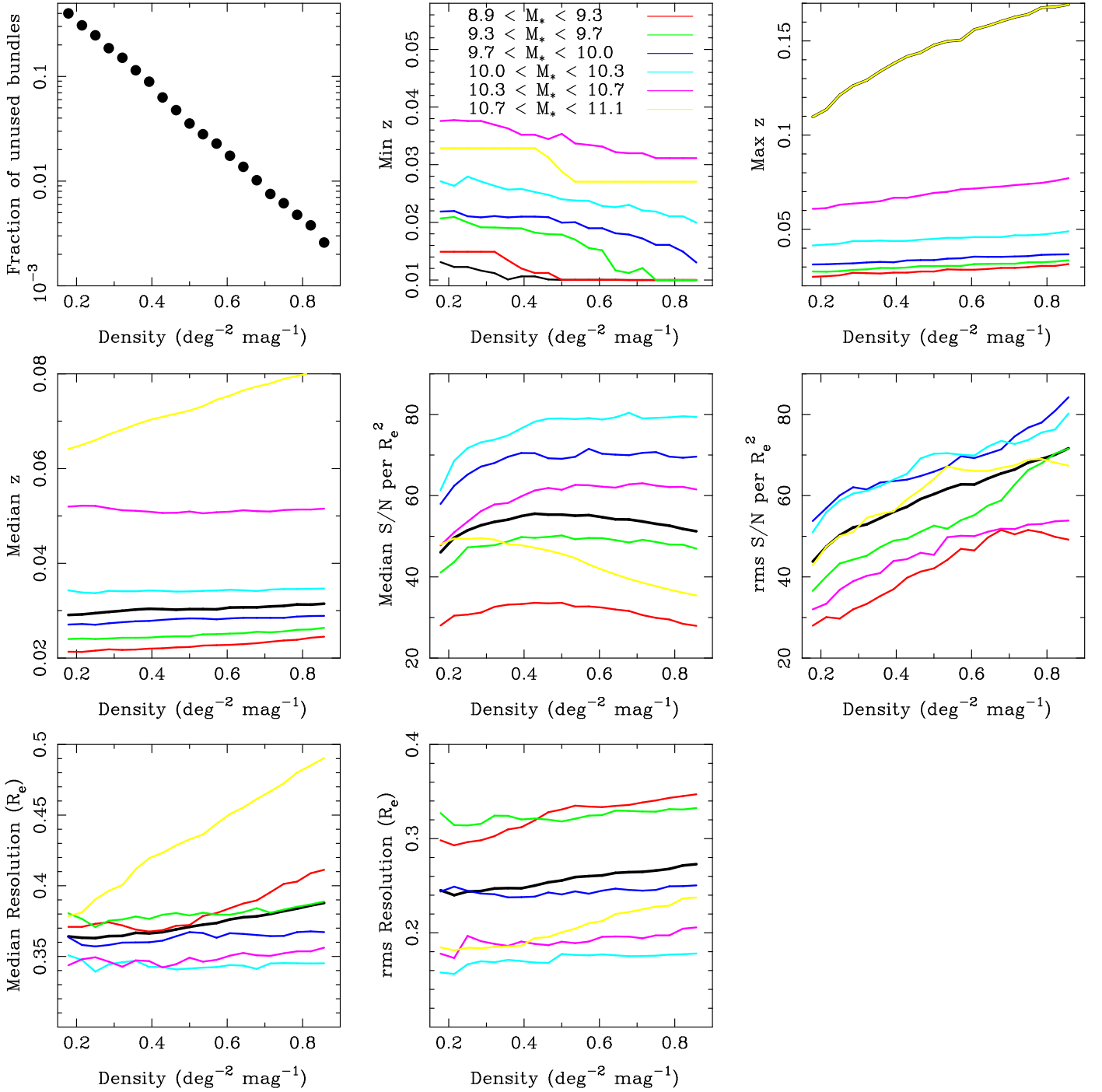
Applying our sample design procedure (Section 3.1) using  $M_i$  as the mass proxy, and defining Primary and Secondary samples that provide coverage to 1.5 and 2.5  $R_e$  for 80% of potential targets results in the  $M_i$  dependent redshift cuts shown in Figures 5 and 6. Applying the IFU size distribution optimization technique (Section 3.2) yields a preferred distribution of 2, 4, 4, 2, 5 for a total of 17 IFUs per plate. It is worth noting here that varying the target density has little effect on the optimal IFU size distribution as long as the target radius (i.e., 1.5 or 2.5  $R_e$ ) and maximum IFU size remain the same. This means that including the Color-Enhanced sample or adjusting our sample selection in the future, for example by changing the relative numbers of Primary and Secondary targets, will not result in a loss of efficiency or introduce a bias from the IFU allocation.

#### 4.3. Primary Sample Properties

Figure 5 shows the properties of the Primary input sample from which actual MaNGA targets will be selected. The top row of panels shows the redshift cuts and the density distribution as a function of both  $M_i$  and  $M_*$ . The center left panel shows the fraction of targets that can be covered to 1.5  $R_e$  with a 127 fiber IFU and the center panel shows the angular size distributions in six equal bins in stellar mass<sup>26</sup> with vertical dashed lines showing the size of the 19 and 127 fiber IFUs. We can see from these panels that the redshift cuts are doing an excellent job of achieving the desired flat  $M_i$  distribution, an approximately flat  $M_*$  distribution and an even angular coverage. All galaxies, irrespective of their mass, have very similar angular size distributions, meaning that they will be covered by the same range in IFU sizes. Of the sample's galaxies, 80.1% have 1.5  $R_e$  smaller than the radius of the 127 fiber IFU with just 3% requiring an IFU smaller than the 19 fiber IFU to reach 1.5  $R_e$ . The introduction of the functional forms for the high- and low-redshift selection cuts results in much smoother cuts at the expense of some minor variation in the density (top middle panel) and coverage fraction (center left panel).

The remaining panels of Figure 5 show distributions of scale and S/N. These are again shown in six mass bins but in these panels the vertical dotted line shows the median for the full

<sup>26</sup> We use  $M_*$  rather than  $M_i$  as it is the underlying physical parameter of interest.

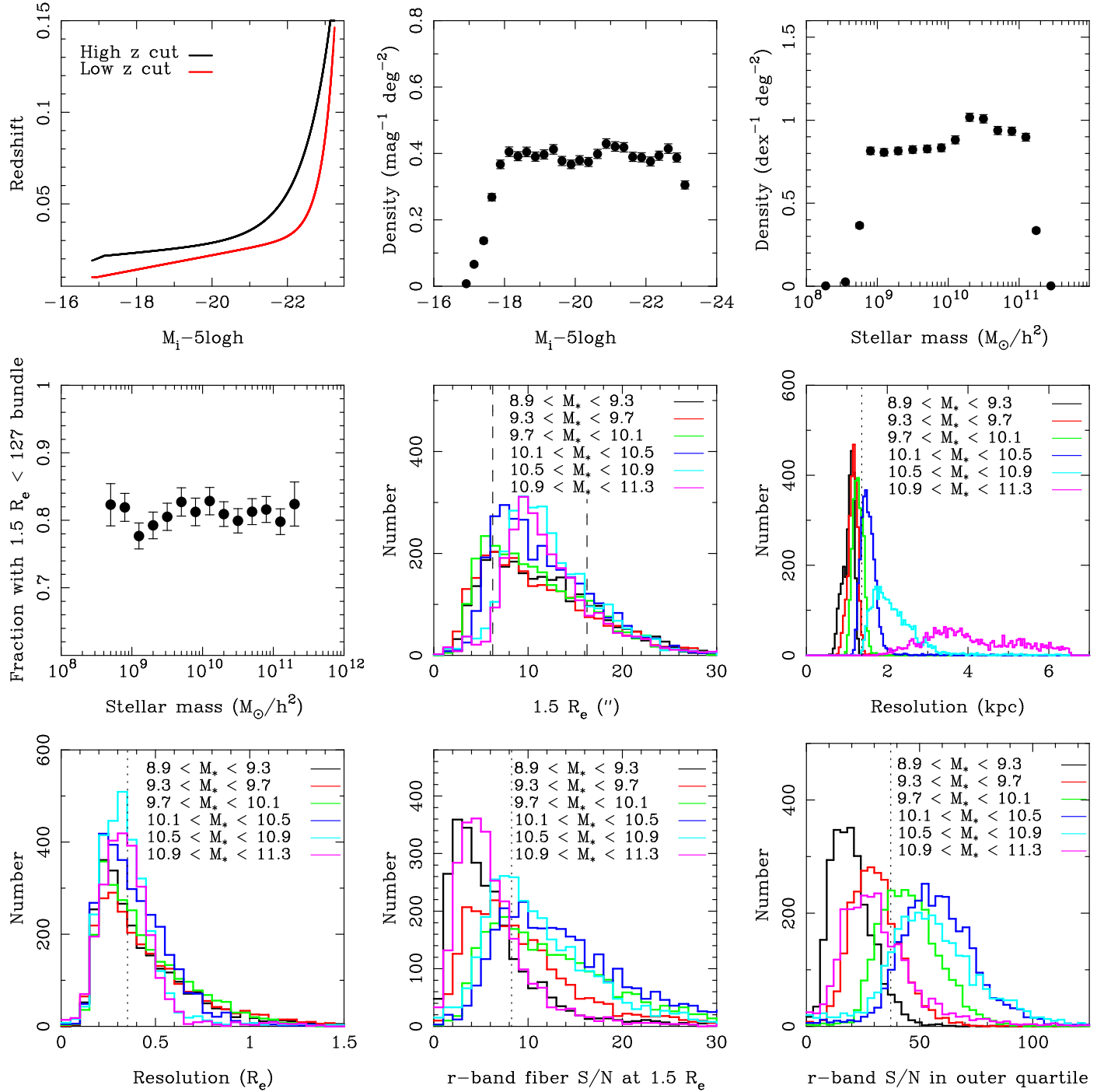


**Figure 4.** Comparison of key properties among samples designed with different target densities per square degree per magnitude. The top left panel shows the fraction of unused IFUs decreases as we increase the number density of potential targets on the sky. The other panels show various properties as a function of the density. In each panel, colored lines indicate different stellar mass bins and the black line indicates the property for the whole sample. See the text for more details.

**Table 3**  
The Fit Parameters for the Functional form (Equation (3)) of the  $M_i - 5 \log h$  Dependent Upper and Lower Redshift Limits Used to Define the Primary and Secondary Samples

Redshift Limit	$A$	$B$	$C$	$D$
Primary lower	-0.056597	-0.0039264	-2.9119	-22.8476
Primary upper	-0.011377	-0.0019220	-1.2924	-22.1610
Secondary lower	-0.056463	-0.0048895	-1.3773	-22.3851
Secondary upper	-0.048010	-0.0046639	-1.3719	-22.3225

sample. The physical resolution in kiloparsecs (center row, right panel) depends strongly on stellar mass, especially at high masses, as a result of the typical redshift increasing as the mass increases. For masses  $< 10^{10} M_{\odot} h^{-2}$  the median resolution is 1.2 kpc, which increases rapidly to a median of 3.85 kpc for the highest mass bin. The median for the full sample is 1.37 kpc. For the most massive galaxies ( $> 10^{10.9} M_{\odot} h^{-2}$ ) there is very little hope of improving the resolution since there are so few of them at low redshifts. For galaxies with  $10^{10} < M_* < 10^{11}$  there are lower redshift galaxies that would yield higher



**Figure 5.** Selection (first three panels) and detailed properties (other panels) of the final Primary sample. Color histograms indicate different stellar mass bins. See the text for detail.

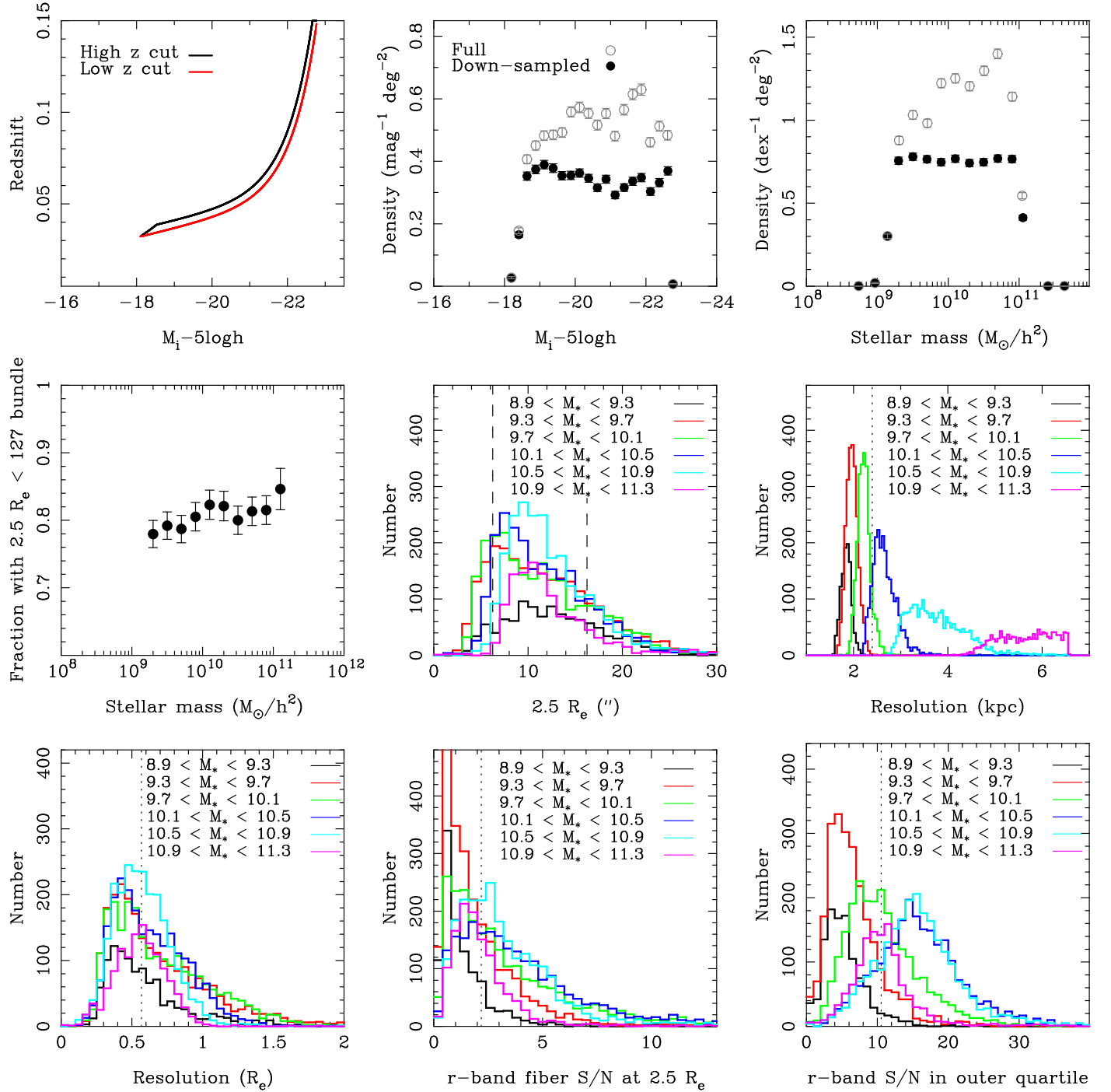
resolutions but these are just too large to be covered to  $1.5 R_e$ . We note that some of these galaxies will be included in an ancillary sample.

When one switches to assessing resolution in terms of  $R_e$  (bottom left panel) then the sample produces distributions that are almost independent of stellar mass. The median resolution for the full sample is  $0.35 R_e$ .

The bottom center and right panels give an indication of the  $r$ -band S/N we can expect to achieve in a typical 3 hr exposure and how it depends on stellar mass. The bottom center panel shows the  $r$ -band S/N per fiber at  $1.5 R_e$  and the bottom right

panel the total S/N in an elliptical annulus (following the ellipticity of the galaxy) covering the outer quartile of the IFU. Typically this annulus will cover  $1.35$ – $1.8 R_e$ . The S/N is lowest for the lowest mass galaxies and increases with mass until the highest masses, where it again decreases. These trends are simply the result of the intrinsic surface brightness distribution of the galaxy population. The medians of the S/N for the full sample at  $1.5 R_e$  are  $8.3$  per fiber and  $37.3$  in the outer quartile annulus. Note here that the S/N refers to the spectral S/N per pixel ( $10^{-4}$  in log wavelength in Å) in the  $r$ -band.





**Figure 6.** Selection (first three panels) and detailed properties (other panels) of the final Secondary sample. Color histograms indicate different stellar mass bins. See the text for detail.

#### 4.4. Secondary Sample Properties

Figure 6 shows the properties of the Secondary sample in the same manner as for the Primary sample above. The  $M_i$  completeness cut Equation (4) is clearly visible in the top left panel as a diagonal cutoff in the high-redshift selection cut at faint magnitudes. This reduces the overall range in  $M_i$  and hence  $M_*$  sampled compared to the Primary sample, limiting the Secondary sample to  $M_* > 2 \times 10^9$ . The other two plots on the top row of the figure again show the density of targets as a function of  $M_i$  and  $M_*$  but unlike the Primary sample there are now two density distributions on each plot. The open symbols

show the density for the full Secondary sample and the filled symbols after it has been down-sampled to a target density half of that of the Primary sample. Our intention was that this down-sampling would be purely random but in error we continued to use an earlier routine that down-sampled to produce a density exactly flat with stellar mass, as can be seen in the figure. The effect of this error is to add an additional weak dependence on stellar mass to the Secondary selection, which needs to be accounted for in any statistical analysis of the MaNGA sample alongside the other selection criteria (see Section 6.1). Since the required correction is well determined and barely more

complex than was already required, and because we had already observed for approximately two years when this error was discovered, we decided that overall it was simpler to continue with this mass-dependent down-sampling of the Secondary sample for the full duration of the survey.

As for the Primary sample, these redshift cuts once again meet our IFU coverage target with 80.7% of the galaxies having  $2.5 R_e$  less than the radius of the 127 fiber IFU, and with just 1% having  $2.5 R_e$  smaller than the 19 fiber IFU. Again the angular size distributions are largely independent of stellar mass.

To achieve such coverage one must select a sample at higher redshift than the Primary sample. This has obvious consequences for both the resolution and S/N. The median resolution is a factor of 1.7 poorer for the Secondary sample compared to the Primary sample, although this is partly due to the Secondary sample being restricted to higher masses as a result of the completeness limits. The S/N at  $2.5 R_e$  (the edge of the IFU) is low with a median S/N of just 2.3 per fiber and 11.4 in the outer quartile annulus. The aim of this sample is not to study continuum properties at  $2.5 R_e$  but those of emission lines, which will naturally have much higher S/N. Even so, the expected S/N in the outer annuli should be sufficient for some absorption line studies, and can be increased further by stacking multiple galaxies.

#### 4.5. Color-Enhanced and Primary+ Samples

While there are many attractive reasons to simply select a sample that is flat in  $M_i$ , the resulting sample is dominated by red galaxies at high masses and blue galaxies at low masses (see Figure 8). A number of MaNGA's primary science drivers concentrate on either red or blue (early or late type) galaxies, and we wish to study how their properties depend on stellar (or dynamical) mass. To improve the statistics for such samples we have designed a Color-Enhanced supplement, which increases the number of galaxies in underpopulated regions of the color (NUV  $- i$ ) versus magnitude ( $M_i$ ) plane. We chose to use NUV  $- i$  rather than another color combination because of the wide dynamic range in color it provides. Even though the NUV fluxes have larger absolute errors than, say, the SDSS  $g$ -band, the much larger dynamic range means that galaxies can still be better separated than if we had used  $g - i$ , for example. The Color-Enhanced supplement is designed to be observed to  $1.5 R_e$  and will be combined with the Primary sample to form the Primary+ sample.

We construct the Color-Enhanced supplement by considering regions of the color–magnitude plane for which the number density of galaxies is  $< 36\%$  of the peak density, which occurs at the luminous end of the red sequence. We calculate these number densities over regions 0.1 mag wide in  $M_i$  and 0.2 mag wide in NUV  $- i$  across a range in absolute magnitude ( $-16.5 > M_i - 5 \log h > -23.5$ ) and color ( $1.6 < \text{NUV} - i < 6.4$ ). When considering the bluest and reddest color bins, we also include galaxies with NUV  $- i < 1.6$  and NUV  $- i > 6.2$ , respectively. For each bin in the color–magnitude plane we expand the redshift limits as defined by the Primary sample to include more galaxies using the following procedure:

1. We expand the redshift limits until the number density for the Primary+ sample in that bin is 36% of the peak density, or the number density is increased by a factor of

three, or we hit the limiting redshift range of the parent catalog ( $0.01 < z < 0.15$ ).

2. If fewer than 50% of the newly added Color-Enhanced galaxies in the bin have  $1.5 R_e$  lying within our IFU size range we reduce the redshift limits of that bin until  $\geq 50$  do have  $1.5 R_e$  lying within our IFU size range, or until half of the candidate Color-Enhanced galaxies have been removed.

These criteria strike a balance between increasing the number density and ensuring coverage to the target radius.

As for the Primary and Secondary samples we always apply a minimum redshift limit of  $z = 0.01$  and a maximum redshift limit below the  $M_i$  completeness limit (Equation (4)).

The demographics of the full Primary+ sample are illustrated in Figure 7. The top row shows from left to right the distribution in the NUV- $i$  versus  $M_i$  plane of the Primary, Color-Enhanced, and Primary+ (the Primary plus the Color-Enhanced) samples. The Color-Enhanced selection is mainly adding galaxies in the *green valley* and in the faint end of the red sequence and the bright end of the blue cloud. The distribution of the Color-Enhanced supplement in the redshift- $M_i$  plane, along with that of the Primary sample is shown in the bottom left panel and the number density of the Primary, Color-Enhanced, and Primary+ samples in the bottom middle panel. The bottom right panel shows that the median angular size of the galaxies in units of fiber diameter, with the error bars showing the 20th and 80th percentiles. The Color-Enhanced supplement typically adds galaxies with a smaller angular size (as a result of their higher redshift) than in the Primary sample at low masses and galaxies of a similar size, but with a slightly larger spread at higher masses. This actually results in a slight increase in the fraction of galaxies in the Primary+ sample that can be covered by the largest IFUs to  $1.5 R_e$  but a slight decrease in the average spatial resolution.

Figure 8 shows the performance of the Primary+ sample in flattening the color–mass distribution of galaxies. The top panel shows the color–mass distribution along with two lines designed to split the sample into red-sequence, green valley, and blue cloud galaxies. The bottom panel shows how the fraction of each of these three galaxy classes depends of stellar mass for both the Primary and Primary+ samples. While there are still more low-mass blue galaxies and high-mass red galaxies in the Primary+ sample, the trends of red and blue fraction with stellar mass have been flattened by the addition of the Color-Enhanced supplement and the fraction of *green valley* galaxies increased.

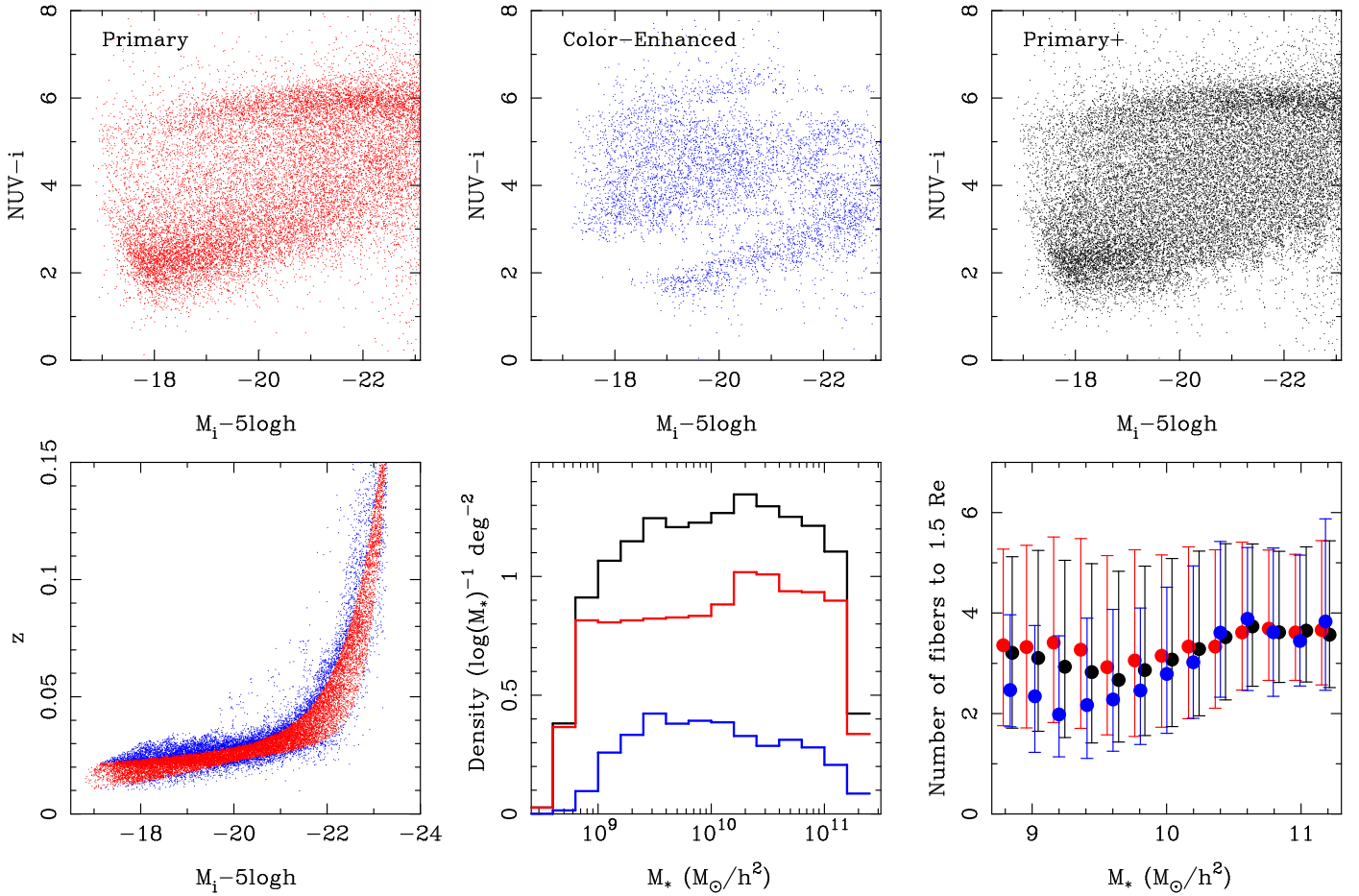
## 5. Tiling the Survey

In order to execute the survey we must decide how we will cover the area spanned by the targeting catalog and allocate IFUs to potential targets. We refer to this process as tiling.

### 5.1. Pointing Strategy

We first discuss how to divide the full area of the input samples into  $7 \text{ deg}^2$  circular tiles, each representing the footprint of a single plate. The degree to which these tiles are allowed to overlap, or indeed repeat, represents a trade-off between efficiency and completeness.

The left panel of Figure 9 shows the distribution of galaxies in our Primary+ and Secondary samples over the full SDSS DR7 footprint, along with an example tiling designed to have

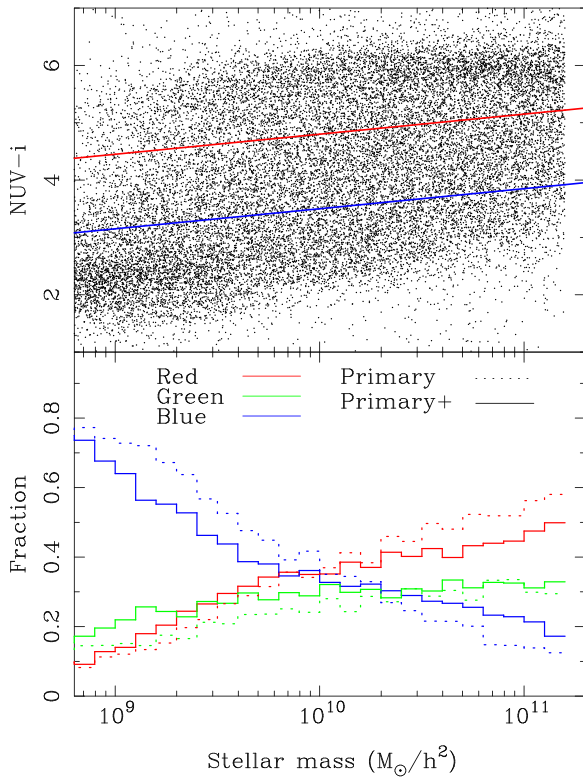


**Figure 7.** Design of the Color-Enhanced supplement. In all panels the Primary sample is shown in red, the Color-Enhanced supplement in blue and the combination (Primary+ sample) in black. The top row shows from left to right the  $NUV - i$  vs.  $i$  distribution for the Primary, the Color-Enhanced, and combination of the two, respectively. The Color-Enhanced supplement fills in the regions of this plane that are sparsely populated in the Primary sample. The bottom left panel shows the distribution in the  $M_i$ -redshift plane for the Primary and Primary+ samples. The Primary+ sample includes galaxies at both higher and low redshifts than the Primary sample. The bottom center panel shows the density distribution of the two samples and the combination of the two, as a function of stellar mass. The roughly flat dependence of density on mass is maintained. The bottom right panel shows the median angular size of the galaxies (in units of fiber diameter) for the two samples and their combination. The error bars show the 20th and 80th percentiles. At high masses we are typically adding large galaxies with the opposite being true at small masses. The overall size distribution remains roughly unchanged when adding the Color-Enhanced supplement to make the Primary+ sample.

non-overlapping tiles. Looking at this figure it is apparent that the number of targets per tile varies considerably; the mean number of potential Primary+ and Secondary MaNGA targets is 27.0 per tile, with an rms of 15.9, a minimum of 2, and a maximum of 233.

One of our requirements is that the selection will be unbiased with respect to environment. This means that we should not choose to ignore a region where we have fewer galaxies than IFUs. It also means that we need to overlap tiles in denser regions on the sky to achieve similar completeness in dense environment as in voids. The right panel of Figure 9 shows the results of our adaptive tiling routine. In this scheme we use a “gravitational” method to assign multiple overlapping plates to the densest regions. The method starts with an evenly distributed overlapping tiling of  $\sim 2000$  tiles. The target galaxies then “pull” tiles as an  $1/r^2$  law, with the mass of already assigned targets set to zero. The velocity with which the plates are allowed to move is damped to prevent oscillation of the tiles and a mild repulsive force between the plates is included. This has the effect of pulling plates toward the over dense regions and away from the voids. Currently only plates with  $>7$  targets are included in the final tiling.

This adaptive tiling scheme is highly effective at increasing our coverage of the densest regions without a significant loss of either IFU allocation efficiency or the ability to assign the correct-size IFU to each target. The IFU allocation efficiency is 97.8% (which increases to 98.5% when the ancillary targets are included; see Section 5.4) and the completeness within the tiled footprint is 87% (which decreases to 85% when the ancillary targets are included), which compares well with the non-overlapping tiling, which has a 93% efficiency and a 60% completeness. In the adaptive scheme, 93% of the galaxies are allocated IFUs matching the target size and 77% are allocated IFUs  $\geq 1.5$  or  $2.5 R_e$  compared to 96.7% and 78% for the non-overlapping tiling. Surprisingly, we see that the overlapping tiling has an increased allocation efficiency but this is simply the result of the removal of tiles that overlap the edge of the imaging region. It also produces a slight decrease in the allocation of IFUs with the desired size. This decrease is caused by the increased completeness, resulting in a reduction in the average number of available IFUs on each tile, making it harder to find galaxies with target sizes matching the available IFU sizes.



**Figure 8.** Top: the distribution of the Primary+ sample in the  $NUV - i$  stellar mass plane. The red and blue lines are used in the bottom panel to divide galaxies into red, green, and blue. Bottom: the fraction of red, green and blue galaxies as a function of stellar mass for the Primary (dotted) and Primary+ (solid) samples. The inclusion of the Color-Enhanced supplement in the Primary+ sample flattens these distributions and increases the numbers of rarer galaxies.

As well as providing a more even sampling of dense environments, our adaptive (overlapping) tiling scheme also enables us to improve observing efficiency. Because we have more total tiles over our area, we have more tiles in the LST regions with fewer overall tiles than in the non-overlapping case, and so can observe the easier tiles in those regions, i.e., ones that are at lower airmass and take less exposure time to complete. Our survey simulations suggest that the adaptive tiling will result in an increase of almost 10% in the total number of plates that we can complete during the survey.

### 5.2. Quality Control and Visual Inspection

Before allocating IFUs to potential targets, we undertook a set of visual inspections of the target catalog to make sure the photometry, and hence the selection of the targets, was reliable. We inspected all galaxies where there was an indication that there may be an issue with the photometry or redshift. Specifically, where the Sérsic and Elliptical Petrosian  $R_e$  differed by a factor of three or more, the Elliptical Petrosian  $R_e > 20''$ , the photometric and spectroscopic (i.e., redshift) centers differed by more than  $1''$ , or the redshift source was not the SDSS. This is a total of 4292 galaxies. For each of these we looked at the imaging, light profiles, and spectra for all of these targets. Galaxies were flagged as bad where the photometry had clearly and significantly failed, due to, e.g., bad imaging, bad deblending with a nearby bright star or galaxy, or a catastrophic background subtraction issue. Galaxies were also flagged where the redshift allocated clearly corresponded to a different

galaxy. Finally we provided a new center if the catalog center was obviously incorrect, mainly due to the presence of a foreground star, bright sub-region, or strong dust lanes.

To check that our preselection was catching the vast majority of issues, we inspected 500 random targets not already flagged as bad, finding no major photometry or centering issues. However, since we do not wish to waste an IFU as the survey proceeds we inspect the targets on each tile before we use that tile to drill any plate that will be observed, again flagging galaxies in the same way. Overall, through the inspection done to date we have flagged 1227 galaxies as bad and recentered 1189, each amounting to  $<3\%$  of the total targets.

Finally we flag the 983 targets lying within  $25''$  of a bright star ( $r < 14$  mag) using the APASS DR8 (Henden et al. 2012) and Tycho 2 (Høg et al. 2000) star catalogs. No targets flagged as bad or close to a bright star are eligible for IFU allocation.

### 5.3. IFU Allocation

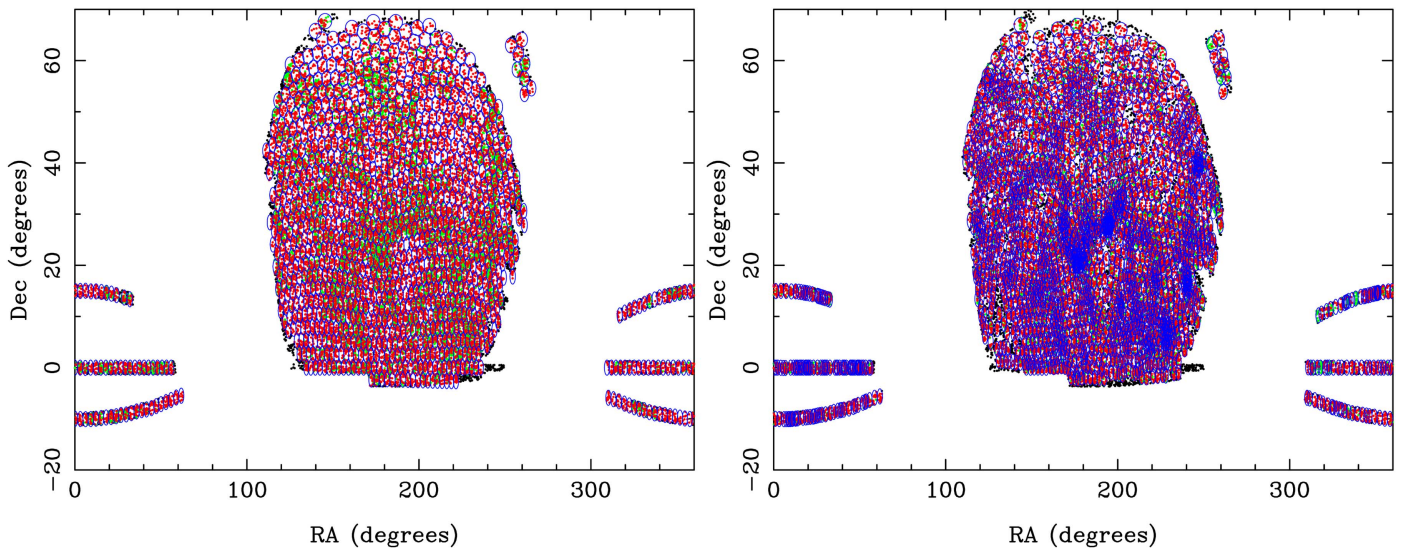
Once we have broken the survey area into tiles, the next step is allocating individual IFUs to galaxies in each tile. Our method for this procedure is designed to maximize the allocation of IFUs to galaxies of the appropriate size. It proceeds as follows:

1. All galaxies within a given tile are selected.
2. Galaxies that collide with the center post (within  $150''$ ) are removed and one of every colliding pair (within  $120''$ ) selected at random is removed.
3. For each IFU size, say 19, galaxies that require a 19 fiber IFU to reach their target radius (e.g.,  $1.5$  or  $2.5 R_e$ ) are selected. Galaxies with a target radius smaller than the 19 fiber IFU are included in the 19 fiber IFU allocation (galaxies larger than the 127 fiber IFU are likewise assigned to the 127 fiber IFU).
4. All available 19 fiber IFUs are assigned to these galaxies.
5. If there are fewer galaxies that need 19 fiber IFUs than available 19 fiber IFUs, the remaining IFUs are put into a pool to be assigned later.
6. This process is repeated for the remaining IFU sizes.
7. The unassigned IFUs are then allocated to the remaining galaxies closest in target size, beginning with the largest galaxies and the largest remaining IFUs, and then working downward.

Galaxies from all the main samples are treated identically (i.e., none are preferred) except those Secondary galaxies that have been removed by down-sampling to get the correct relative number of Primary and Secondary targets. These randomly removed Secondaries will only be allocated an IFU if there are no available targets from another sample.

This method will produce a sample that has an angular size distribution that matches as closely as possible the IFU size distribution. Since we have optimized the IFU size distribution (see Section 3.2) to match the actual galaxy target size distribution we expect that any bias introduced will be small. However, we expect some bias to be introduced given that we are using a 2, 4, 4, 2, 5 IFU size distribution when 2, 4, 3, 2, 5 was slightly more optimal and that we only have a small integer number of each size IFU, which is highly unlikely to match the desired IFU size distribution perfectly. However, any such bias can very easily be corrected by comparing the relative number of IFUs of a given size required for the sample to the number available (see Section 6.2 for details of these weights). We





**Figure 9.** Illustration of two potential tiling strategies. The plot on the left shows a non-overlapping tiling and the plot on the right shows an adaptive overlapping tiling designed to produce a more even completeness irrespective of target density. The black points show the positions of galaxies in our final Primary+ and Secondary targeting catalogs. The blue ellipses show the outlines of the tiles. The green points show those galaxies that lie within the footprint of these tiles and the red points those galaxies that were assigned IFUs.

demonstrate that the bias is indeed small and corrected by appropriate weights in Figure 10 by determining the fraction of galaxies that are allocated an IFU as a function of stellar mass and size ( $R_e$ ).

Figure 10 shows the results of applying our allocation method to the combined Primary+ and Secondary samples using the adaptive tiling presented in Section 5, and our optimized 2, 4, 4, 2, 5 IFU size distribution along with an IFU allocation weight (see Section 6.2). In all panels we show the fraction of galaxies allocated an IFU in blue, and the fraction allocated an IFU of the correct size, which is where the IFU is either greater than or equal to the target size (the smaller value between 1.5 or 2.5  $R_e$  and the radius of 127 fiber IFU), in red. The raw fractions are shown with the lighter colored open points and the fractions corrected by the IFU allocation weights by the darker filled points. The left panels show these fractions as a function of stellar mass, the right panels as a function of  $R_e$ , and the top and bottom panels for Primary and Secondary samples respectively.

We can see that almost 90% of all galaxies covered by our tiling are allocated an IFU and that there is no trend apparent with stellar mass for either the raw or corrected fractions. In the raw counts there is a weak trend with  $R_e$  particularly within the Secondary sample, such that small galaxies are slightly more likely to be allocated an IFU than larger ones. This results from the slightly non-optimal match between our IFU size distribution and that required for the final sample, which has been made a little worse by small changes in our sample definition, such as switching from  $M_*$  to  $M_i$  adding in the Color-Enhanced supplement and using functional forms to describe our selection cuts, since the 2, 4, 4, 2, 5 IFU size distribution was defined. However, the inclusion of the weights completely corrects for this bias leaving no trend with physical size.

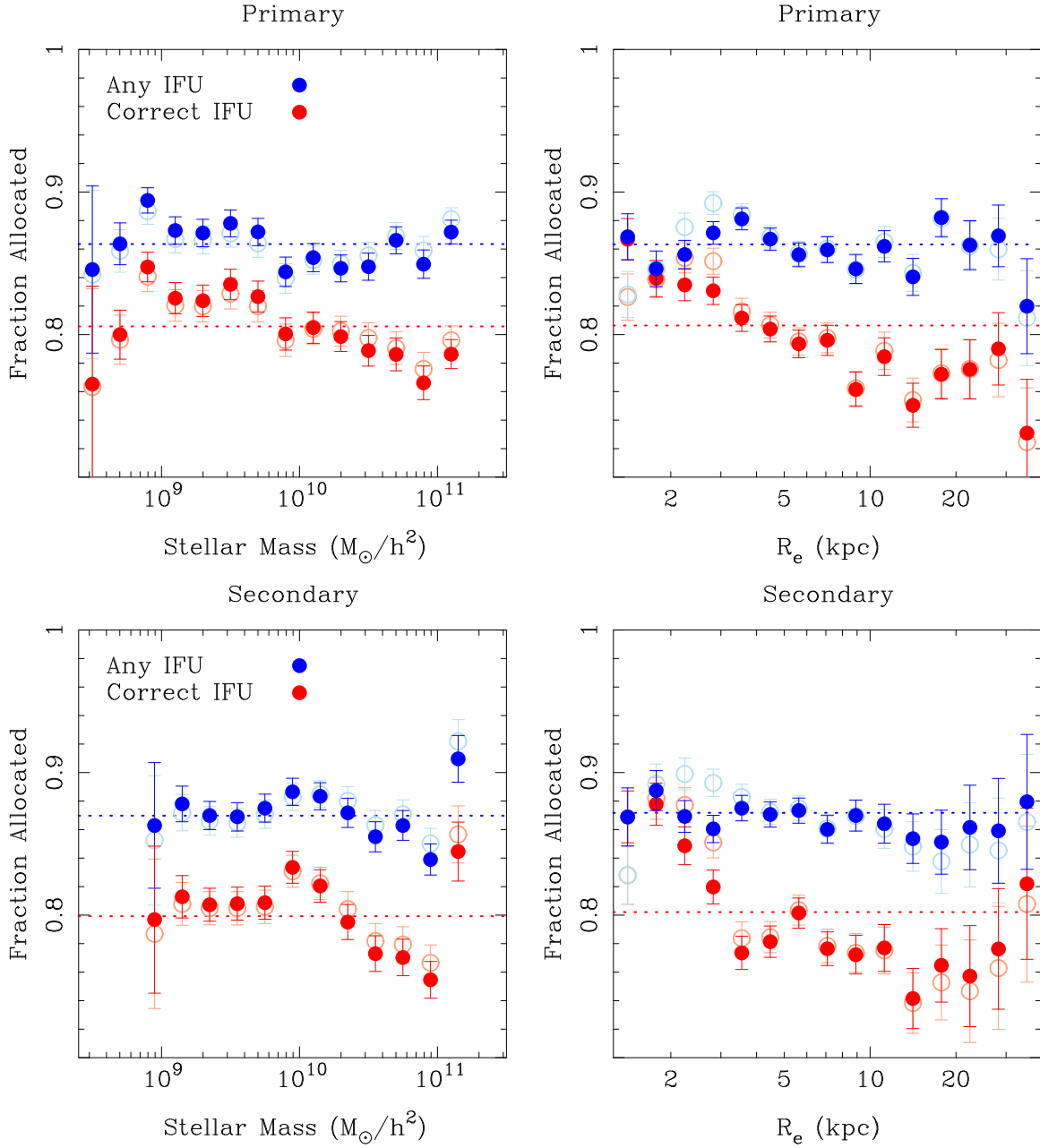
When we turn to look at the galaxies that are covered by an IFU that is actually larger than their target radius, we see the trend with size becomes stronger in both samples. This likely results from the necessity to allocate IFUs to galaxies that do not exactly match the required size when there are not enough galaxies on a tile matching the available IFU sizes. As

described above, those initially unallocated IFUs are still assigned to the remaining galaxy targets on the tile. Since, for a given mass, intrinsically small galaxies will also have small angular sizes, they are easier to cover to their target radius with whatever IFU is available. Whereas an intrinsically large galaxy may only be covered by one of the larger IFUs.

In Figure 11 we show how the allocation efficiency depends on environment. In addition to improving efficiency, the main goal of our adaptive tiling scheme is to ensure that the fraction of galaxies allocated IFUs is independent of environment. This is of particular concern in the densest environments where most of the galaxies remain untargeted with just a single plate. We choose to define environment using the Yang et al. SDSS DR7 group catalog (Yang et al. 2007), splitting our target galaxies into centrals and satellites and by group luminosity. We also indicate the expected halo mass one might associate with group luminosity using the determinations from the catalog.

The left panels of Figure 11 show the distribution of the numbers of all the selected Primary and Secondary sample galaxies (solid histograms) and those that were allocated an IFU (dotted histograms). Since we have an approximately flat distribution in stellar mass we end up with an approximately flat distribution in halo mass for the central galaxies, with slightly extended tails to high and low halo mass. The extended tail is particularly noticeable at the high halo mass end where the scatter in the halo mass hosting a given stellar mass galaxy is larger. Satellites occupy groups of higher luminosity or mass than centrals with the same stellar mass and so as expected the distribution is offset from the centrals.

The right-hand panels show how the fraction of galaxies allocated an IFU depends on group luminosity, again split by central and satellite designation. For the Primary sample there is very little dependence in the allocated fraction below  $10^{11} L_\odot h^{-2}$ , but above that there is a general trend of increasing allocation fraction with increasing group luminosity, particularly at the highest luminosities. This trend is caused by only being able to allocate tiles in integer numbers and the removal of excess tiles with fewer than seven allocated targets. For galaxies in lower luminosity groups, which are covered on



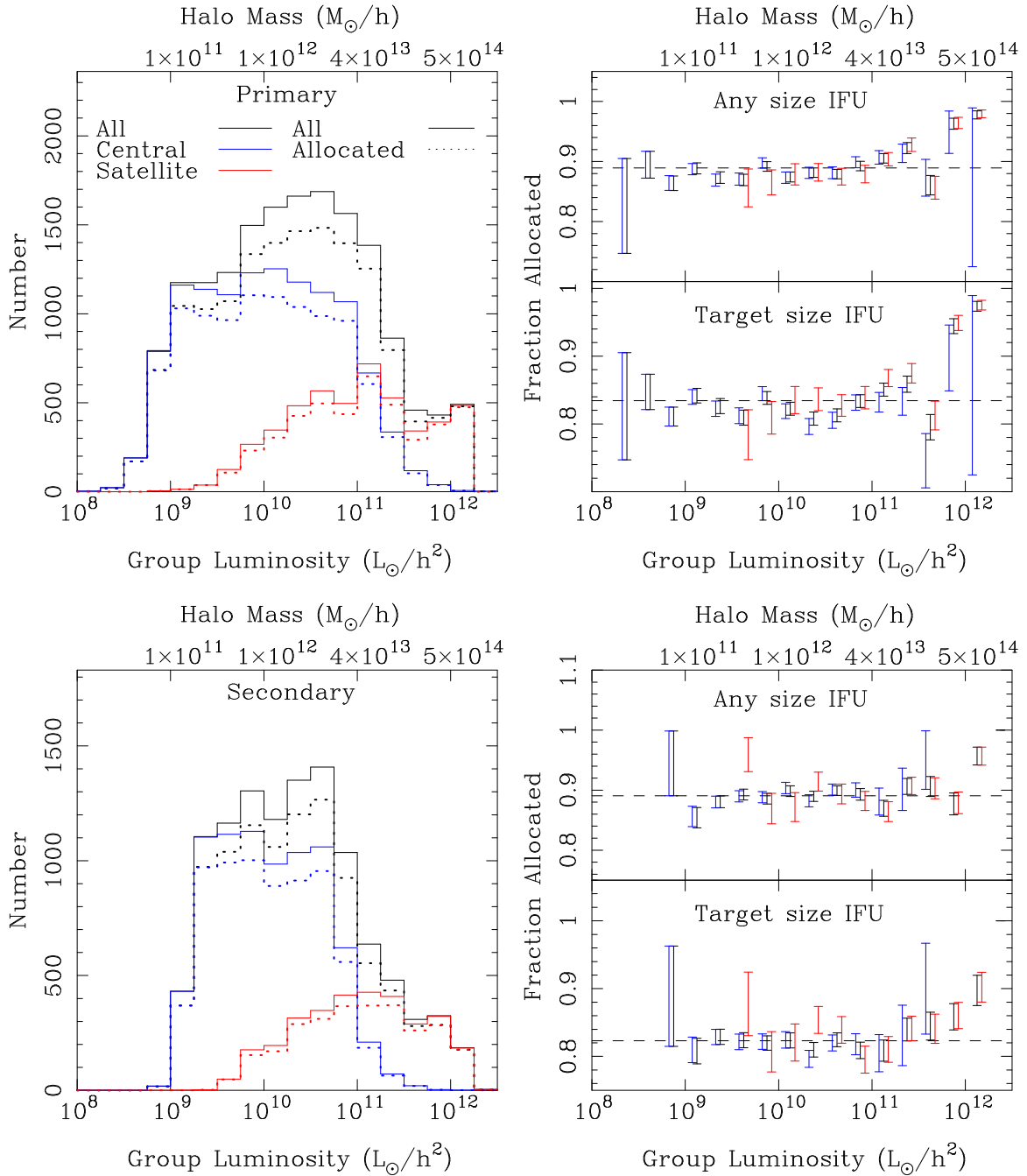
**Figure 10.** Fraction of galaxies allocated an IFU (blue) and the fraction allocated an IFU of the correct size (red) as a function of stellar mass (left) and  $R_e$  (right) for the Primary (top) and Secondary (bottom) samples. The dashed lines show the fraction irrespective of stellar mass or size. The open lighter colored points show the raw fractions, whereas the filled darker points show the fractions corrected using the IFU size allocation bias weight (see text for details).

average by  $\sim 1.8$  tiles, the removal of a tile with just a few allocated targets has a significant impact on completeness in that region. Galaxies in the most luminous groups are typically in the densest regions and are covered by significantly more tiles (an average of  $10.7$  for  $10^{12} L_{\odot} h^{-2}$  groups). As a result, the removal of a single tile with fewer than seven allocated galaxies in such a dense region has a negligible effect on the allocation completeness, leaving the fraction of galaxies allocated an IFU close to  $100\%$ . For the Secondary sample we see almost no dependence of the allocation fraction on group luminosity apart from in the most luminous groups ( $>10^{12} L_{\odot} h^{-2}$ ). The less significant trend simply reflects the smaller number of Secondary sample galaxies and so only the most luminous groups have enough of an effect on the overall sample density to have an impact on the allocation fraction.

Finally, in Table 4 we present the numbers of potential targets and allocated IFUs for our final adaptive tiling of 1800 MaNGA tiles that covers the full DR7 area, after excluding tiles containing fewer than seven targets. We have assumed an IFU size distribution of 2, 4, 4, 2, 5 and the Primary, Color-enhanced, and Secondary samples as described above.

Simulations of the survey observing strategy suggest that MaNGA can complete  $\sim 600$  plates in six years, so simply dividing the numbers in this table by three will give an indication of the size of the final MaNGA samples and show we should reach our goal of 10,000 galaxies for the three main samples.

Over the full tiling of 1800 tiles we are unable to allocate 453 IFUs due to a lack of targets, which is  $1.5\%$  of the available IFUs. These unallocated IFUs will be allocated to



**Figure 11.** Left panels: the distribution in the number of galaxies targeted as a function of group luminosity for the Primary (top) and Secondary (bottom) sample. The solid histograms show all targets and the dotted just those allocated an IFU. We additionally divide the targets into centrals (blue) and satellites (red). Right panels: the fraction of galaxies allocated an IFU of any size (top subpanel) and with a size greater than or equal to the target size (bottom subpanel) as a function of group luminosity for the Primary (top) and Secondary (bottom). All galaxies are shown in black, centrals in blue and satellites in red as before.

repeated observations of galaxies on overlapping tiles as a test of our data, or finally on the rare occasions this is not possible, to galaxies that lie just outside our main selection criteria.

#### 5.4. Ancillary Targets

In order to enhance the broader scientific goals of the MaNGA survey a call for proposals within the collaboration was issued for Ancillary targets that would use a small fraction of the MaNGA IFUs. These programs typically target rarer classes of galaxies that represent important phases in galaxy evolution, but are under-represented in the three main samples

(e.g., luminous AGN, mergers, central galaxies in massive halos) or they target galaxies useful in testing, calibrating, or understanding the main MaNGA data (e.g., high spatial resolution in a nearby galaxy, galaxies with calibrated kinematics). The goal for these programs was to make use of allocated IFUs as well as a small fraction of IFUs that could have been allocated to main sample galaxies.

The IFU allocation for the chosen Ancillary programs was made after the allocation to the three main samples. Each program provided a list of targets ranked by priority as well as an IFU size requirement for each target. In addition to the internal priority within each program, each program was

**Table 4**  
IFU Allocation Results for an Adaptive MaNGA Tiling of 1800 Plates

Input Sample	Targets	Got IFU	$R_{\text{IFU}} \geq \text{Target Size}$	$R_{\text{IFU}} \geq R_s$
Combined	38162	29204 (0.765)	27111 (0.928)	22413 (0.767)
Combined DS	32772	27806 (0.848)	25917 (0.932)	21395 (0.769)
Primary+	21661	18334 (0.846)	17190 (0.938)	14286 (0.779)
Primary	16231	13747 (0.847)	12867 (0.936)	10514 (0.765)
Secondary	16533	10898 (0.659)	9947 (0.913)	8149 (0.748)
Secondary DS	11143	9500 (0.853)	8753 (0.921)	7131 (0.751)
Color-En	5430	4587 (0.845)	4323 (0.942)	3772 (0.822)

**Note.** Numbers in parentheses are fractions.  $R_s$  is  $1.5 R_e$  for the Primary sample and  $2.5 R_e$  for the Secondary sample. The target IFU size may still be less than  $R_s$  for large galaxies. DS refers to the Secondary sample after it has been down-sampled to its target density. These results include the allocation of IFUs to ancillary targets.

prioritized based on how it was ranked by the internal Ancillary TAC. Throughout the ancillary allocation procedure IFUs were allocated in order of these priorities.

The ancillary allocation proceeded in three passes. On the first pass we allocate as many of the unallocated IFUs that match the target size as possible. On the second pass we re-allocate IFUs that have already been allocated to Secondary galaxies that did not pass the down-sampling cut (identified as  $\text{RANFLAG} = 0$  in the catalog) that have the required size. On the third pass we re-allocate IFUs that have been allocated to main sample targets (i.e., Primary+ and Secondary post down-sampling) up to a certain fraction. Since some of the Ancillary targets were already in the main samples and just required a different IFU size the final step is to re-allocate the IFUs removed from these targets to a new Primary+ or Secondary sample target of a suitable size if one is available.

In total we allocate 5.1% of the available IFUs to ancillaries, 3% to entirely new ancillary targets and 2.1% to already allocated targets in the main samples that required a larger IFU size. As a result the fraction of Primary+ and Secondary sample galaxies allocated an IFU reduces from 86.6% to 84.8%. The fraction of unallocated IFUs decreases from 2.3% to 1.5%.

Detailed descriptions of the individual ancillary programs are available on the SDSS data release pages from DR13 onwards.

## 6. Considerations When Using the MaNGA Samples

The MaNGA sample is selected in a somewhat different manner to typical galaxy samples, which are often either magnitude or volume limited, and so there are certain considerations to take into account when using it for science. In this section we describe how to approach using the sample and some things that should be avoided.

### 6.1. Volume Weights

As is the case for many galaxy samples the MaNGA sample is not volume limited, although that is by design rather than observational necessity. We have chosen to flatten the number density distribution as a function of  $M_i$  so that we have as many high luminosity galaxies as we have low luminosity galaxies. As already discussed, this will produce a fairly even S/N on measured quantities as a function of  $M_i$  (or stellar mass), but could (and most likely will) introduce a bias for any analysis as a function of anything other than  $M_i$  unless corrections back to a volume-limited sample are made. Even in a narrow stellar mass bin the distribution of galaxy mass to light ratio (or star formation rate or metallicity, etc.) will not be the same as the

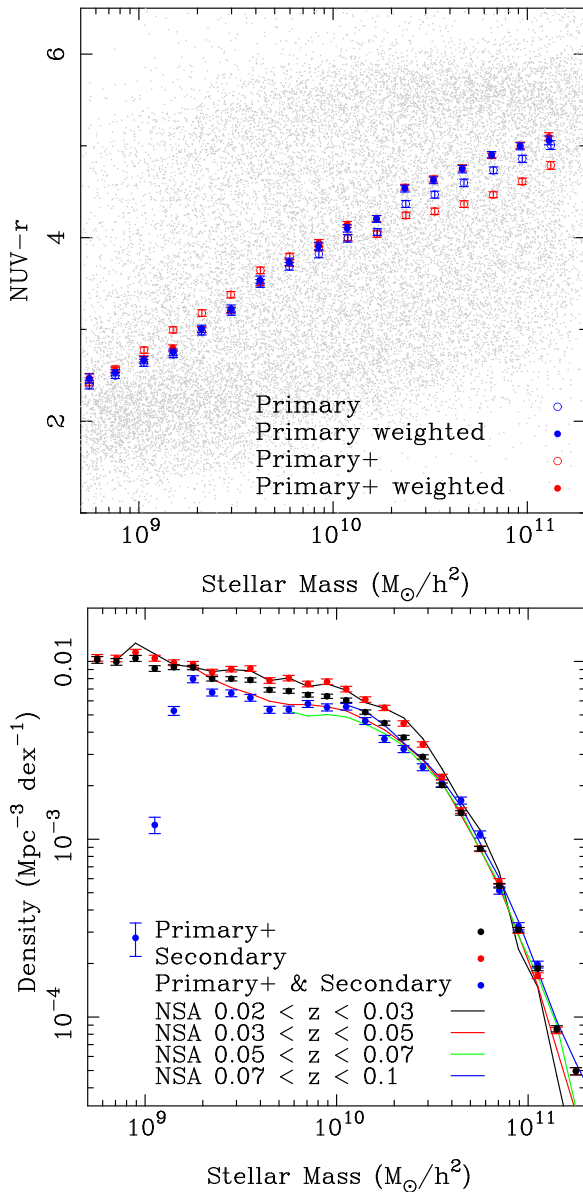
distribution in that mass bin for a volume-limited sample, even if the mass bin were very narrow.

Fortunately the MaNGA selection is designed to make the calculation of the required volume weights very simple since we know for every galaxy the redshift range over which it could have been selected. At a given  $M_i$  (or  $M_i$  and  $\text{NUV} - r$  color for the Primary+ sample) the sample is effectively volume limited in that all galaxies with  $z_{\min}(M_i) < z < z_{\max}(M_i)$  are targeted irrespective of their other properties. However, that volume varies with  $M_i$ . Therefore, in any analysis of the properties of MaNGA galaxies as a function of anything other than  $M_i$  we must correct for this varying selection volume,  $V_s(M_i)$  (the volume with  $z_{\min}(M_i) < z < z_{\max}(M_i)$ ). The simplest approach is just to correct the galaxies back to a volume-limited sample by applying a weight ( $W$ ) to each galaxy in any calculation, such that  $W = V_f/V_s$  where  $V_f$  is an arbitrary fiducial volume.

For each galaxy in MaNGA we provide ZMIN, ZMAX, SZMIN, SZMAX, EZMIN, and EZMAX, which are the minimum and maximum redshift each galaxy could have been observed over for the Primary, Secondary, and Primary+ samples, respectively. So for a given galaxy one can just convert the appropriate (given which sample it is in)  $z_{\min}$  and  $z_{\max}$  to the volume ( $V_s$ ) and the weight is  $1/V_s$ .

An additional complication arises if one is using the Secondary sample on its own or in conjunction with the Primary or Primary+ samples. The Secondary sample was down-sampled to the appropriate surface density so that the 2/3 to 1/3 Primary+ to Secondary sample split would be achieved, with an average sampling rate of  $\sim 67.1\%$ . However, we allow unallocated IFUs that cannot be allocated to any other targets to be allocated to Secondaries not included in the down-sampling, which effectively changes the Secondary sampling rate to 76.9%. These additional galaxies will most likely be on plates with low surface densities of targets and so could be biased toward lower density regions. If you were concerned about such a bias, the safest approach is to ignore these additional galaxies and so restrict the Secondary sample to galaxies with  $\text{RANFLAG} = 1$ . To first order you could then multiply the Secondary weights by  $1/0.671$  to reflect the down-sampling rate or if you are not concerned about this potential bias then you can use all the Secondaries multiplying the weight by  $1/0.769$ . However, since the down-sampling rate depends weakly on stellar mass (see Section 4.4), ideally the down-sampling correction to the weight needs to be made on a galaxy-by-galaxy basis. A full description of how to calculate weights for each sample and combinations of samples is given





**Figure 12.** Volume weight corrections. Top: the  $\text{NUV} - r$  colors of the Primary+ galaxies as a function of stellar mass. The open blue and red points show the median colors for the Primary and Primary+ samples, respectively. The filled red and blue points show the median colors for the two samples after the  $V_{\text{max}}$  weights have been applied, effectively correcting the median colors back to those of a volume-limited sample. Bottom: the stellar mass functions of the Primary+ sample (red points), the Secondary sample (blue points), and their combination (black points) compared to stellar mass functions calculated using the full NSA catalog in a series of redshift ranges (lines).

in [Appendix](#) and these weights will be made available with DR14.

Please note that you should never use the Color-Enhanced supplement on its own for statistical population studies. There are regions of color-magnitude space that are not populated in the Color-Enhanced supplement. If you want to include the Color-Enhanced supplement you should use the Primary+ sample (Primary + Color-Enhanced) where all regions of our nominal color-magnitude space are sampled. EZMIN and EZMAX are defined for all galaxies in the Primary+ sample.

Figure 12 illustrates the application and importance of the volume weights. In the top panel we plot the  $\text{NUV} - r$  color of the Primary+ sample as a function of stellar mass along with

the mean  $\text{NUV} - r$  for the Primary and Primary+ samples with and without the volume weights applied. The volume weights used here are calculated relative to a reference volume such that the mean weight is equal to one. Without the volume weights the Primary sample shows bluer mean colors than with the weights, with the difference becoming larger at larger masses. When flattening in  $M_i$  as we move from low to high galaxy luminosity we are increasing the selection volume, and hence relative number of galaxies, compared to a volume-limited sample. As a result for a given stellar mass you select more blue galaxies, which have a smaller mass-to-light ratio and are hence more luminous, than you would for a volume-limited sample. The magnitude of this effect becomes larger at higher masses, corresponding to brighter  $M_i$  where the luminosity function is steep and so the volume change required to flatten the number density becomes rapid. Turning to the Primary+ sample we can see a much flatter trend in mean  $\text{NUV} - r$  than the Primary sample by design, since we’ve preferentially filled in under-represented regions of the color-magnitude plane with the Color-Enhanced supplement. When the volume weights are applied, the Primary+ mean colors exactly match those of the Primary sample since they are now both weighted back to a volume-limited sample.

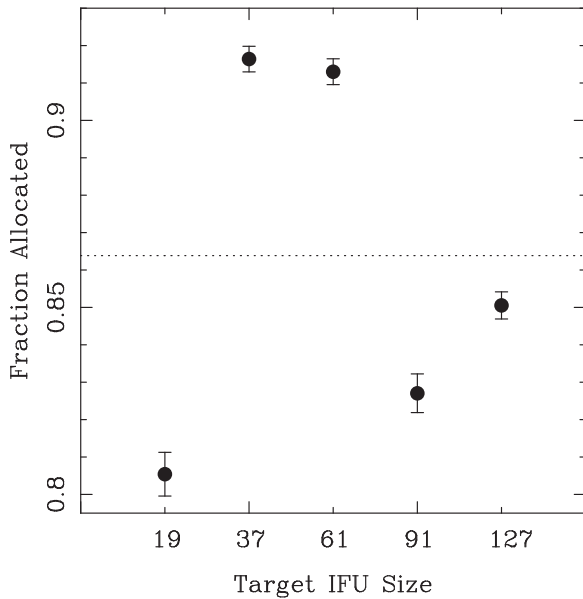
The bottom panel of Figure 12 shows the reconstructed stellar mass function of the three main samples (points) compared to mass functions calculated using the full extended NSA sample in different redshift intervals (lines). Here we have used a  $1/V_{\text{max}}$  weight with  $V_{\text{max}}$  calculated using the area of the survey<sup>27</sup> and the selection redshift limits appropriate for each sample. The NSA mass functions are also calculated using a  $1/V_{\text{max}}$  weight where  $V_{\text{max}}$  is determined using the stellar mass completeness limit that results from the magnitude-limited nature of the NSA. The figure nicely illustrates that the volume weights modify the approximately flat stellar mass density distributions seen in Figures 5–7 to the expected Schechter function shape of the mass function. We can also see good agreement between the mass functions of the individual samples and those calculated using the full NSA in the redshift ranges that overlap well with a given sample.

Finally Figure 12 demonstrates the advantage of being able to select a sample with a constant number density as a function of  $M_i$ . Whereas the errors on the NSA mass functions depend strongly on mass (there would be an even stronger dependence with a volume-limited sample rather than the magnitude-limited NSA) they are essentially constant for the MaNGA sample.

## 6.2. IFU Allocation Weights

To most efficiently cover our targets to the desired radii, we select galaxies that match in size to the available IFU sizes on each tile. For several reasons (see Sections 3.2 and 5.3 for a discussion of these), and despite careful optimization, the MaNGA IFU size distribution does not exactly match the distribution required for our final samples. This slight mismatch is demonstrated in Figure 13 where we show the fraction of IFUs allocated as a function of the target IFU size for each galaxy. The allocation fraction varies from almost 0.85 to 0.95 with targets requiring the 19 fiber IFU being the least likely to

<sup>27</sup> In this figure we have used the full targeting catalog. To make such a plot from observed MaNGA data one must use the appropriate area and completeness corresponding to the set of observed plates.



**Figure 13.** Fraction of galaxies allocated an IFU as a function of target IFU size for all galaxies in the Primary+ and Secondary (after down-sampling) samples. The overall fraction is shown as the dotted line. There is a small variation in the allocation fraction with target IFU size as a result of the IFU size distribution not being perfectly optimal.

be allocated an IFU and targets requiring the 37 fiber IFU being the most likely. Although the variation is small, less than 5% from the mean, it will introduce a slight angular size selection bias, which translates to a physical size bias (see Figure 10) and hence a surface mass density bias, etc.

However, once again it is very simple to correct this bias with a weight. We define the allocation weight for a galaxy with a given IFU target size  $S_{\text{IFU}}$  as  $W_e(S_{\text{IFU}}) = A(S_{\text{IFU}})/A_{\text{all}}$  where  $A(S_{\text{IFU}})$  is the allocation fraction for that IFU target size and  $A_{\text{all}}$  is the allocation fraction for all targets, i.e., the ratio of the points to the line in Figure 13. Table 5 gives the weights for each IFU target size for a combined sample of the Primary+ and Secondary (after down-sampling, RANFLAG = 1) targets. We combine the samples since the IFU allocation procedure is blind to the them and so the correction should be the same for them all. We do not include in this calculation the Secondary targets that are initially excluded by down-sampling since they are allocated only when there are no other targets remaining on a tile and so will have a different size bias, as well as the likely environment bias we discussed in Section 6.1.

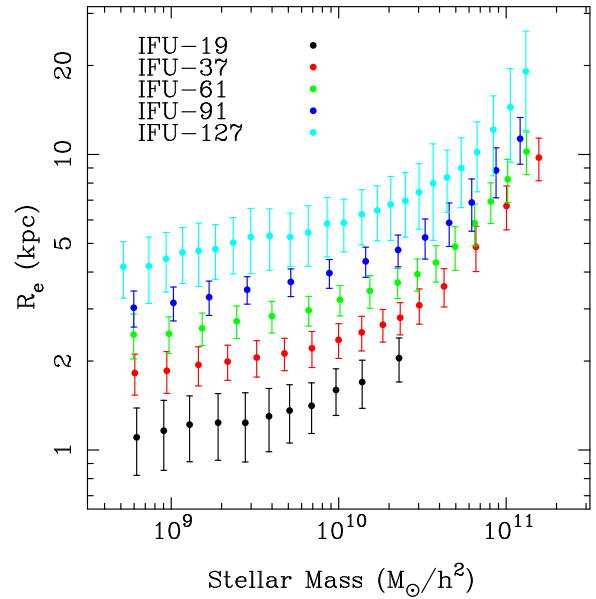
Figure 10 shows that applying these weights removes any bias in the allocation fraction with  $R_e$ .

### 6.3. Selecting by IFU Size—Do Not Do It!

In many cases the precision with which one could make a certain measurement from the MaNGA data will depend on the number of spatial resolution elements there are covering a galaxy. Since MaNGA uses a range of IFU sizes the number of spatial resolution elements varies by a factor of 6.7 between the largest and smallest MaNGA IFUs. It may well then be tempting to use only the largest IFUs for a given science analysis. However, this will introduce a significant bias in the sample selected. For a given  $M_i$  (or mass) the redshift range over which the MaNGA targets are selected is narrow, corresponding to a small range in the spatial resolution being

**Table 5**  
Weights Required to Correct for the IFU Size-dependent Allocation Completeness

IFU Target Size	Allocation Weight
19	1.075
37	0.945
61	0.949
91	1.047
127	1.018



**Figure 14.** Mean  $R_e$  as a function of stellar mass for Primary sample galaxies split by IFU target size. Larger IFUs are typically assigned to intrinsically larger galaxies of a given mass. This could introduce a significant bias if only a subset of the IFUs were used for a given analysis.

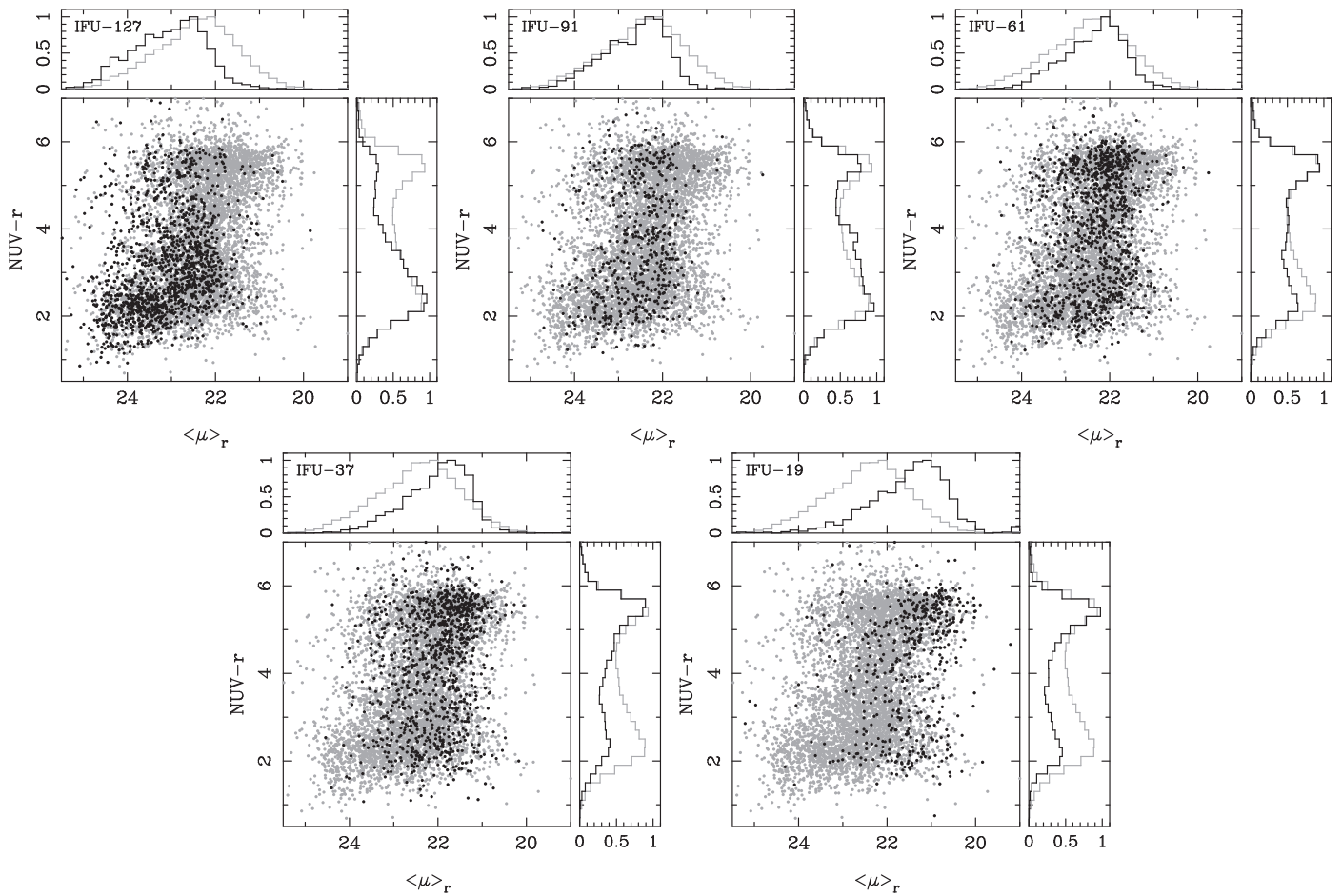
probed. This means that larger IFUs are typically being assigned to intrinsically larger galaxies.

Figures 14 and 15 illustrate the correlation between IFU size and intrinsic galaxy size within MaNGA and the consequences of considering only certain IFU sizes. Figure 14 shows the mean  $R_e$  as a function of stellar mass for Primary sample galaxies split by IFU target size. As expected, at every mass the galaxies requiring a 19 fiber IFU are on average significantly smaller than those requiring a 127 fiber IFU by up to a factor of four. The consequences of this are shown in Figure 15 where we show the distribution of the Primary galaxies in the NUV –  $r$  color versus  $r$ -band surface brightness plane for galaxies targeted with each size of IFU. The largest IFUs are biased toward low surface brightness and hence toward bluer galaxies, whereas the opposite is true for the smallest IFUs with a continuous trend between the two extremes for the intermediate-sized IFUs.

As a result we would advise that great caution is taken to ensure that the aims of any analysis that uses only a subset of the IFU sizes is independent of such biases.

### 6.4. Selecting by Inclination

Selecting subsamples by inclination will likely be desirable for many science analyses using MaNGA data, whether it is favoring or avoiding edge- or face-on galaxies. Such subsamples introduce no bias to the distribution of the properties of the subsamples as long as the standard volume



**Figure 15.** Distribution of Primary sample galaxies in the  $\text{NUV} - r$  vs.  $r$ -band surface brightness plane. The full sample is shown as the light gray points and histograms in every panel, whereas the black points and histograms in each panel are for a different IFU size. The large IFUs are more likely to be allocated to low-surface-brightness blue galaxies, whereas the small IFUs are more likely to be allocated to high-surface-brightness red galaxies.

weights are used. One might worry that the increase in extinction as galaxies become viewed more edge-on could introduce a bias and indeed without the weights there is a small inclination bias in the sample, such that there are fewer edge-on galaxies than face-on galaxies at a given stellar mass than one would find in a volume-limited sample. This bias results from the extra extinction of an edge-on galaxy results in a fainter  $M_i$  at a given  $M_*$ , meaning it would only be selected over a lower and narrower redshift range and hence a smaller volume. The same would happen if we selected by extinction specifically, or by color, or SFR, and is simply the result of using  $M_i$  in the selection rather than mass. All of these biases are removed by using the volume weights. The exception to this comes at low masses, which we discuss next.

### 6.5. The Lowest Mass Galaxies

The only circumstances where the volume weights may not completely remove selection bias as intended is at the lowest masses. Since our selection is based on  $M_i$  (or  $M_i$  and color), at a fixed stellar mass some fraction of the galaxies may not be in our sample at all. For instance red galaxies, either because they are old, dusty, or inclined. This is simply the familiar stellar mass incompleteness present in all samples selected using luminosity. The color-dependent  $M_i$  that we use at faint magnitudes (Section 4.1) greatly reduces the magnitude of this effect, but it will be an issue for the reddest galaxies. We

estimate that we are virtually complete for  $M_* > 5 \times 10^8 M_\odot h^{-2}$  for the Primary and Primary+ samples and  $M_* > 2 \times 10^9 M_\odot h^{-2}$  for the Secondary sample. We note that we are most likely still incomplete for the most inclined galaxies at these stellar mass limits, where the  $i$ -band extinction may be higher than for face-on galaxies by close to one magnitude. As such, analyses making use of only the most inclined systems should approach these completeness limits with caution.

### 6.6. Fiber/IFU Collisions

As already discussed, the overall fraction of target galaxies allocated an IFU is  $\sim 84\%$  with only a very weak dependence on size, stellar mass, or halo mass. However, as a result of collisions between IFUs in MaNGA and between spectroscopic fibers in the original SDSS I/II spectroscopy (Blanton et al. 2003), there is a significant decrease in completeness for pairs of galaxies with small separations  $< 120''$  ( $8h^{-1}$  kpc to  $115h^{-1}$  kpc over the MaNGA redshift range). The fraction of MaNGA targets that are in a collision (pairs with separations  $< 120''$ ) is 8% and for these targets the fraction allocated an IFU is decreased to 71%.

In addition there is a further incompleteness within the MaNGA target catalog as a result of spectroscopic incompleteness in the extended NSA. This incompleteness is mainly the result of fiber collisions within the original SDSS spectroscopy,



which occur at pair separations  $<55''$ , and results in an overall spectroscopic completeness of the extended NSA of  $\sim 98\%$ . We estimate that for MaNGA targets with pairs separated by  $<55''$  the spectroscopic completeness of the target catalog is reduced at most to 69%. This is a lower limit since to make this calculation we have assumed that galaxies with missing redshifts have the same redshift as their nearest neighbor. If that were not the case the likelihood of both galaxies in a close pair being targeted by MaNGA is reduced. The final completeness within  $55''$  will then be further reduced as a result of the MaNGA IFU collisions to  $\sim 47\%$ .

While these reductions in completeness for close pairs may seem large, it is important to remember that this is affecting a very small fraction of the full galaxy population and so for most purposes making global corrections for incompleteness (one for the extended NSA and one for MaNGA IFU allocation) should be sufficient. However, if one is studying close pairs and the study requires an absolute normalization, such as would be needed for merger fractions as a function of separation or very small-scale clustering, further corrections should be made (e.g., Guo et al. 2012; Patton et al. 2016).

## 7. Summary and Conclusions

Given the bounds of the SDSS telescope and the BOSS spectrograph, the MaNGA sample has been designed alongside the IFU size distribution and allocation strategy to efficiently produce a sample of approximately 10,000 galaxies with a relatively even sampling of color–magnitude space and galaxy environment, along with radial coverage to either  $1.5 R_e$  or  $2.5 R_e$ . This selection has been achieved using simple criteria based purely on  $i$ -band absolute magnitude (and  $\text{NUV} - i$  color for a small subset of targets) and redshift.

We have found that to most efficiently target such a sample requires the following:

1. A range in IFU sizes, to account for the intrinsic variation of galaxy size even at fixed mass and redshift.
2. The selection of more luminous or more massive galaxies at higher redshift, to produce an angular size distribution that is largely independent of mass.
3. The density of targets, which is determined by the volume probed at any given  $M_i$ , must be sufficiently large so as to provide enough targets as to be able to efficiently allocate IFUs of the correct size, thus minimizing unused IFUs, and not too large so as to produce too wide a range in galaxy angular size.
4. Within those constraints, the targets should be at as low a redshift as possible to maximize S/N and spatial resolution.

Following our optimization we select an IFU size distribution with  $2 \times 19$  fiber,  $4 \times 37$  fiber,  $4 \times 61$  fiber,  $2 \times 91$  fiber, and  $5 \times 127$  fiber IFUs, and design two samples, the Primary+ sample with coverage to  $1.5 R_e$  and the Secondary sample with coverage to  $2.5 R_e$ . This dual-sample approach was designed to allow exploration of the dark matter-dominated regime and poorly studied outskirts of galaxies, while still also producing a sample that is spatially well sampled in the high S/N inner regions of galaxies that contain the bulk of the stellar mass.

The Primary+ sample, making up two-thirds of the targets, is made up of the Primary sample selected purely by  $M_i$  and redshift and the Color-Enhanced supplement, which includes

an additional selection in  $\text{NUV} - i$  color and fills in low-density regions of color–magnitude space. The median  $r$ -band per fiber S/N at  $1.5 R_e$  for the Primary sample is expected to be 8.3 with little dependence on galaxy mass other than for the lowest mass galaxies. The median spatial resolution is  $0.35 R_e$  and 1.37 kpc (1.2 kpc for  $M_* < 10^{10} M_\odot h^{-2}$ , 3.8 kpc for  $M_* > 10^{11} M_\odot h^{-2}$ ).

The Secondary sample, making up one-third of the targets, is, like the Primary sample, selected purely on  $M_i$  and redshift but targets higher redshift galaxies in order to provide coverage to  $2.5 R_e$  with the same size IFUs. As a result, both the spatial resolution and S/N are reduced. The median resolution is a factor of 1.7 poorer, and the S/N is just 2.3 per fiber at  $2.5 R_e$ . For this sample, the primary science is expected to come from emission-line studies, where the line S/N will be much higher, or through stacking multiple fibers in single galaxies or from samples of galaxies.

While the even sampling in stellar mass and color greatly enhances the usefulness of the MaNGA sample for studies over the full range of galaxy properties, it does require some care to avoid subtle biases being introduced into any population analysis. However, since the selection is simply defined by a redshift range at a given  $M_i$  (or  $M_i$  and  $\text{NUV} - i$  color), it is straightforward to calculate a volume weight for each galaxy, which corrects back to a volume-limited sample, thus eliminating these biases.

It is clear from the MaNGA data that has already been taken that this sample will achieve its objectives of enabling a wide range of statistical analyses of the spatial distribution of galaxy properties in the local universe, thus providing great insight into the physics of galaxy formation.

The full MaNGA target catalog including many parameters from the NSA and IFU allocation information is publicly available as part of the SDSS data releases from DR13, with the volume weights included from DR14 onward.

A.W. acknowledges support of a Leverhulme Trust Early Career Fellowship.

This research was made possible through the use of the AAVSO Photometric All-Sky Survey (APASS), funded by the Robert Martin Ayers Sciences Fund.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is [www.sdss.org](http://www.sdss.org).

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of



Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

### Appendix Volume Weight Calculation

We define a volume for each galaxy in each of the three samples as the co-moving volume enclosed by the redshift selection limits for that galaxy and the tiled area of the survey ( $7362 \text{ deg}^2$ ),  $V_P$ ,  $V_S$  and  $V_{P+}$  for the Primary, Secondary and Primary+ sample galaxies respectively. We also define a fiducial volume  $V_F = 1 \times 10^6 \text{ Mpc}^3$ , which is approximately the average volume of the MaNGA samples, although this choice is arbitrary.

To correct for the allocation incompleteness we calculate the probability  $P$  that a given galaxy is allocated an IFU. For the Primary and Primary+ samples this is simply the ratio of the number of targets allocated an IFU to the total number that could have been, i.e., that lie on a tile,<sup>28</sup> and is the same for all galaxies in that sample. So

$$P_P = N_{Pa}/N_P \quad (5)$$

and

$$P_{P+} = N_{P+a}/N_{P+} \quad (6)$$

where  $N$  is the number of galaxies, and subscript  $P$  and  $P+$  indicate the Primary and Primary+ samples, and subscript  $a$  indicates only those allocated an IFU.

For the Secondary sample we must also take into account the effect of the stellar mass-dependent down-sampling and so the allocation probability is uniquely defined for each galaxy. Additionally, since in the initial main allocation stage we allocate only to those galaxies that *pass* the down-sampling ( $\text{RANFLAG} = 1$ ) and then allocate any remaining unallocated IFUs to those that do not *pass* ( $\text{RANFLAG} = 0$ ), we must treat the down-sampled and non-down-sampled Secondaries separately, calculating  $P_{Sd}$  and  $P_{Snd}$ , respectively. So we have

$$P_{Sd}(M_*) = P_d(M_*) \times N_{Sda}/N_{Sd} \quad (7)$$

and

$$P_{Snd}(M_*) = [1 - P_d(M_*)] \times N_{Snda}/N_{Sd} \quad (8)$$

where  $P_d(M_*)$  is the probability that a given galaxy is included after the stellar mass-dependent down-sampling. The allocation probability for the full Secondary sample is then simply

$$P_S(M_*) = P_{Sd}(M_*) + P_{Snd}(M_*). \quad (9)$$

We then combine these allocation probabilities with the maximum volumes to produce the weights for each individual

sample as

$$W_P = V_F/(V_P P_P) \quad (10)$$

$$W_S = V_F/(V_S P_S) \quad (11)$$

$$W_{Sd} = V_F/(V_S P_{Sd}) \quad (12)$$

$$W_{P+} = V_F/(V_{P+} P_{P+}) \quad (13)$$

and for combinations of samples as

$$W_{PSd} = V_F/(V_P P_P + V_S P_{Sd})$$

$$W_{P+,Sd} = V_F/(V_{P+,m} P_{P+} + V_S P_{Sd}) \quad (14)$$










$$W_{PS} = V_F/(V_P P_P + V_S P_S) \quad (15)$$

$$W_{P+,S} = V_F/(V_{P+,m} P_{P+} + V_S P_S). \quad (16)$$

Where we are combining the Primary+ and Secondary sample we modify the Primary+ volume ( $V_{P+,m}$ ) so we do not double count in the rare instances where the upper Primary+ redshift limit overlaps with the lower Secondary redshift limit.

Applying these weights to MaNGA targets allocated an IFU within the tiled area of the full MaNGA target catalog will produce a volume-limited sample with a volume of  $1 \times 10^6 \text{ Mpc}^3$ , and so densities are simply  $N/1 \times 10^6 \text{ Mpc}^{-3}$  where  $N$  is the weighted number of galaxies. If only a subset of the galaxies is being used, as will most likely be the case, the volume must be scaled by the ratio of the observed area to the fully tiled area of  $7362 \text{ deg}^2$ .

### ORCID iDs

David A. Wake  <https://orcid.org/0000-0002-6047-1010>  
 Kevin Bundy  <https://orcid.org/0000-0001-9742-3138>  
 Renbin Yan  <https://orcid.org/0000-0003-1025-1711>  
 Michael R. Blanton  <https://orcid.org/0000-0003-1641-6222>  
 Matthew A. Bershady  <https://orcid.org/0000-0002-3131-4374>  
 Niv Drory  <https://orcid.org/0000-0002-7339-3170>  
 David R. Law  <https://orcid.org/0000-0002-9402-186X>  
 Karen Masters  <https://orcid.org/0000-0003-0846-9578>  
 Joel R. Brownstein  <https://orcid.org/0000-0002-8725-1069>

### References

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543  
 Bacon, R., Accardo, M., Adjali, L., et al. 2010, *Proc. SPIE*, **7735**, 773508  
 Bershady, M. A., Verheijen, M. A. W., Swaters, R. A., et al. 2010, *ApJ*, **716**, 198  
 Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, *AJ*, **154**, 28  
 Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, *AJ*, **142**, 31  
 Blanton, M. R., Lin, H., Lupton, R. H., et al. 2003, *AJ*, **125**, 2276  
 Blanton, M. R., & Roweis, S. 2007, *AJ*, **133**, 734  
 Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, *AJ*, **129**, 2562  
 Bryant, J. J., Owers, M. S., Robotham, A. S. G., et al. 2015, *MNRAS*, **447**, 2857  
 Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, *ApJ*, **798**, 7  
 Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, *MNRAS*, **413**, 813  
 Colless, M., Dalton, G., Maddox, S., et al. 2001, *MNRAS*, **328**, 1039  
 Croom, S. M., Lawrence, J. S., Bland-Hawthorn, J., et al. 2012, *MNRAS*, **421**, 872  
 Drory, N., MacDonald, N., Bershady, M. A., et al. 2015, *AJ*, **149**, 77  
 Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, *AJ*, **130**, 2598  
 Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, *AJ*, **131**, 2332  
 Guo, H., Zehavi, I., & Zheng, Z. 2012, *ApJ*, **756**, 127

<sup>28</sup> The completeness will vary as the survey progresses with more and more overlapping tiles being observed with time. As a result, if absolute rather than relative volumes weights are required, this probability and hence the weights should be recalculated for each data release.

- Henden, A. A., Levine, S. E., Terrell, D., Smith, T. C., & Welch, D. 2012, *JAVSO*, **40**, 430
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, *A&A*, **355**, L27
- Huchra, J. P., & Geller, M. J. 1991, in *ASP Conf. Ser. 15, Large-scale Structures and Peculiar Motions in the Universe*, ed. D. W. Latham & L. A. N. da Costa (San Francisco, CA: ASP), 143
- Jones, D. H., Read, M. A., Saunders, W., et al. 2009, *MNRAS*, **399**, 683
- Martin, C., Moore, A., Morrissey, P., et al. 2010, *Proc. SPIE*, **7735**, 77350M
- Patton, D. R., Qamar, F. D., Ellison, S. L., et al. 2016, *MNRAS*, **461**, 2589
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, *A&A*, **538**, A8
- Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnell, A. W. 2011, *ApJS*, **196**, 11
- Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, *AJ*, **146**, 32
- Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, *AJ*, **124**, 1810
- Yan, R., Bundy, K., Law, D. R., et al. 2016, *AJ*, **152**, 197
- Yang, X., Mo, H. J., van den Bosch, F. C., et al. 2007, *ApJ*, **671**, 153